Document AN2 WORLD INTELLECT Appl. No. 09/937,484

International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: C12N 15/31, 1/21, C12P 21/02, C07K 14/33, A61K 38/16, 39/08

A1

(11) International Publication Number:

WO 98/07864

(43) International Publication Date:

(GB).

26 February 1998 (26.02.98)

(21) International Application Number:

PCT/GB97/02273

(22) International Filing Date:

22 August 1997 (22.08.97)

(30) Priority Data:

9617671.4 9625996.5

23 August 1996 (23.08.96) GB

13 December 1996 (13.12.96) GB

(74) Agent: SCHLICH, George, William; Mathys & Squire, 100 Gray's Inn Road, London WC1X 8AL (GB).

Conrad, Padraig [GB/GB]; Microbiological Research

Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB).

FOSTER, Keith, Alan [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG

(60) Parent Application or Grant

(63) Related by Continuation US

Filed on

08/782,893 (CIP) 27 December 1996 (27.12.96)

(71) Applicants (for all designated States except US): MICRO-BIOLOGICAL RESEARCH AUTHORITY CAMR (CEN-TRE FOR APPLIED MICROBIOLOGY & RESEARCH) [GB/GB]; Porton Down, Salisbury, Wiltshire SP4 0JG (GB). THE SPEYWOOD LABORATORY LIMITED [GB/GB]: 14 Kensington Square, London W8 5HH (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): SHONE, Clifford, Charles [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). QUINN, (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD,

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: RECOMBINANT TOXIN FRAGMENTS

(57) Abstract

A polypeptide has first and second domains which enable the polypeptide to be translocated into a target cell or which increase the solubility of the polypeptide, or both, and further enable the polypeptide to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis. The polypeptide thus combines useful properties of a clostridial toxin, such as a botulinum or tetanus toxin, without the toxicity associated with the natural molecule. The polypeptide can also contain a third domain that targets it to a specific cell, rendering the polypeptide useful in inhibition of exocytosis in target cells. Fusion proteins comprising the polypeptide, nucleic acids encoding the polypeptide and methods of making the polypeptide are also provided. Controlled activation of the polypeptide is possible and the polypeptide can be incorporated into vaccines and toxin assays.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

			•				
AL AM AT AU AZ BBB BE BF BG BJ BRY CAF CCI CCM CCI CCZ	Albania Armenia Austria Austria Australia Azerbaijan Bosnia and Herzegovina Barbados Belgium Burkina Faso Bulgaria Benin Brazil Belarus Canada Central African Republic Congo Switzerland Côte d'Ivoire Cameroon China Cuba Czech Republic	ES FI FR GA GB GE GH GN GR HU IE IL IS IT JP KE KG KP KZ	Spain Finland France Gabon United Kingdom Georgia Ghana Guinea Greece Hungary Ireland Israel Iceland Italy Japan Kenya Kyrgyzstan Democratic People's Republic of Korea Republic of Korea Kazakstan Saint Lucia	LS LT LU LY MC MD MG MK ML MN MR MW MX NE NL NO NZ PL PT RO RU	Lesotho Lithuania Luxembourg Latvia Monaco Republic of Moldova Madagascar The former Yugoslav Republic of Macedonia Mali Mongolia Mauritania Malawi Mexico Niger Netherlands Norway New Zealand Poland Portugal Romania Russian Federation	SI SK SN SZ TD TG TJ TM TR TT UA UG US UZ VN YU ZW	Slovenia Slovakia Scregat Swaziland Chad Togo Tajikistan Turkmenistan Turkey Trinidad and Tobago Ukraine Uganda United States of America Uzbekistan Viet Nam Yugoslavin Zimbabwe
CN CU	China Cuba	KZ	Republic of Korea Kazakstan	PT RO	Portugal		

RECOMBINANT TOXIN FRAGMENTS

This invention relates to recombinant toxin fragments, to DNA encoding these fragments and to their uses such as in a vaccine and for *in vitro* and *in vivo* purposes.

The clostridial neurotoxins are potent inhibitors of calcium-dependent neurotransmitter secretion in neuronal cells. They are currently considered to mediate this activity through a specific endoproteolytic cleavage of at least one of three vesicle or pre-synaptic membrane associated proteins VAMP, syntaxin or SNAP-25 which are central to the vesicle docking and membrane fusion events of neurotransmitter secretion. The neuronal cell targeting of tetanus and botulinum neurotoxins is considered to be a receptor mediated event following which the toxins become internalised and subsequently traffic to the appropriate intracellular compartment where they effect their endopeptidase activity.

The clostridial neurotoxins share a common architecture of a catalytic L-chain (LC, ca 50 kDa) disulphide linked to a receptor binding and translocating H-chain (HC, ca 100 kDa). The HC polypeptide is considered to comprise all or part of two distinct functional domains. The carboxy-terminal half of the HC (ca 50 kDa), termed the $H_{\rm C}$ domain, is involved in the high affinity, neurospecific binding of the neurotoxin to cell surface receptors on the target neuron, whilst the amino-terminal half, termed the $H_{\rm N}$ domain (ca 50 kDa), is considered to mediate the translocation of at least some portion of the neurotoxin across cellular membranes such that the functional activity of the LC is expressed within the target cell. The $H_{\rm N}$ domain also has the property, under conditions of low pH, of forming ion-permeable channels in lipid membranes, this may in some manner relate to its translocation function.

For botulinum neurotoxin type A (BoNT/A) these domains are considered to reside within amino acid residues 872-1296 for the H_c , amino acid residues 449-871 for the H_N and residues 1-448 for the LC. Digestion with trypsin effectively degrades the H_C domain of the BoNT/A to generate a non-toxic fragment designated LH_N ,

which is no longer able to bind to and enter neurons (Fig. 1). The LH_N fragment so produced also has the property of enhanced solubility compared to both the parent holotoxin and the isolated LC.

It is therefore possible to provide functional definitions of the domains within the neurotoxin molecule, as follows:

- (A) clostridial neurotoxin light chain:
- -a metalloprotease exhibiting high substrate specificity for vesicle and/or plasma membrane associated proteins involved in the exocytotic process. In particular, it cleaves one or more of SNAP-25, VAMP (synaptobrevin / cellubrevin) and syntaxin.
- (B) clostridial neurotoxin heavy chain H_N domain:
- -a portion of the heavy chain which enables translocation of that portion of the neurotoxin molecule such that a functional expression of light chain activity occurs within a target cell.
- -the domain responsible for translocation of the endopeptidase activity, following binding of neurotoxin to its specific cell surface receptor via the binding domain, into the target cell.
- -the domain responsible for formation of ion-permeable pores in lipid membranes under conditions of low pH.
- -the domain responsible for increasing the solubility of the entire polypeptide compared to the solubility of light chain alone.
- (C) clostridial neurotoxin heavy chain H_c domain.
- -a portion of the heavy chain which is responsible for binding of the native

holotoxin to cell surface receptor(s) involved in the intoxicating action of clostridial toxin prior to internalisation of the toxin into the cell.

The identity of the cellular recognition markers for these toxins is currently not understood and no specific receptor species have yet been identified although Kozaki et al. have reported that synaptotagmin may be the receptor for botulinum neurotoxin type B. It is probable that each of the neurotoxins has a different receptor.

It is desirable to have positive controls for toxin assays, to develop clostridial toxin vaccines and to develop therapeutic agents incorporating desirable properties of clostridial toxin.

However, due to its extreme toxicity, the handling of native toxin is hazardous.

The present invention seeks to overcome or at least ameliorate problems associated with production and handling of clostridial toxin.

Accordingly, the invention provides a polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to neuronal exocytosis and wherein said second domain is adapted (i) to translocate the polypeptide into the cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into the cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of any clostridial neurotoxin precursor that can be converted into toxin by proteolytic action. Accordingly, the invention may thus provide a single polypeptide chain containing a domain equivalent to a clostridial toxin light chain and a domain providing the functional aspects of the H_N of a clostridial toxin heavy chain, whilst lacking the functional aspects of a clostridial toxin H_C domain.

For the purposes of the invention, the functional property or properties of the H_N of a clostridial toxin heavy chain that are required to be exhibited by the second domain of the polypeptide of the invention are either (i) translocation of the polypeptide into a cell, or (ii) increasing solubility of the polypeptide compared to solubility of the first domain on its own or (iii) both (i) and (ii). References hereafter to a H_N domain or to the functions of a H_N domain are references to this property or properties. The second domain is not required to exhibit other properties of the H_N domain of a clostridial toxin heavy chain.

A polypeptide of the invention can thus be soluble but lack the translocation function of a native toxin-this is of use in providing an immunogen for vaccinating or assisting to vaccinate an individual against challenge by toxin. In a specific embodiment of the invention described in an example below a polypeptide designated LH₄₂₃/A elicited neutralising antibodies against type A neurotoxin. A polypeptide of the invention can likewise thus be relatively insoluble but retain the translocation function of a native toxin - this is of use if solubility is imparted to a composition made up of that polypeptide and one or more other components by one or more of said other components.

The first domain of the polypeptide of the invention cleaves one or more vesicle or plasma-membrane associated proteins essential to the specific cellular process of exocytosis, and cleavage of these proteins results in inhibition of exocytosis, typically in a non-cytotoxic manner. The cell or cells affected are not restricted to a particular type or subgroup but can include both neuronal and non-neuronal cells. The activity of clostridial neurotoxins in inhibiting exocytosis has, indeed, been observed almost universally in eukaryotic cells expressing a relevant cell surface receptor, including such diverse cells as from Aplysia (sea slug), Drosophila (fruit fly) and mammalian nerve cells, and the activity of the first domain is to be understood as including a corresponding range of cells.

The polypeptide of the invention may be obtained by expression of a recombinant nucleic acid, preferably a DNA, and is a single polypeptide, that is to say not

cleaved into separate light and heavy chain domains. The polypeptide is thus available in convenient and large quantities using recombinant techniques.

In a polypeptide according to the invention, said first domain preferably comprises a clostridial toxin light chain or a fragment or variant of a clostridial toxin light chain. The fragment is optionally an N-terminal, or C-terminal fragment of the light chain, or is an internal fragment, so long as it substantially retains the ability to cleave the vesicle or plasma-membrane associated protein essential to exocytosis. The minimal domains necessary for the activity of the light chain of clostridial toxins are described in J. Biol. Chem., Vol.267, No. 21, July 1992, pages 14721-14729. The variant has a different peptide sequence from the light chain or from the fragment, though it too is capable of cleaving the vesicle or plasma-membrane associated protein. It is conveniently obtained by insertion, deletion and/or substitution of a light chain or fragment thereof. In embodiments of the invention described below a variant sequence comprises (i) an N-terminal extension to a clostridial toxin light chain or fragment (ii) a clostridial toxin light chain or fragment modified by alteration of at least one amino acid (iii) a C-terminal extension to a clostridial toxin light chain or fragment, or (iv) combinations of 2 or more of (i)-(iii).

In further embodiments of the invention, the variant contains an amino acid sequence modified so that (a) there is no protease sensitive region between the LC and H_N components of the polypeptide, or (b) the protease sensitive region is specific for a particular protease. This latter embodiment is of use if it is desired to activate the endopeptidase activity of the light chain in a particular environment or cell. Though, in general, the polypeptides of the invention are activated prior to administration.

The first domain preferably exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin. The clostridial toxin is preferably botulinum toxin or tetanus toxin.

In an embodiment of the invention described in an example below, the toxin light

chain and the portion of the toxin heavy chain are of botulinum toxin type A. In a further embodiment of the invention described in an example below, the toxin light chain and the portion of the toxin heavy chain are of botulinum toxin type B. The polypeptide optionally comprises a light chain or fragment or variant of one toxin type and a heavy chain or fragment or variant of another toxin type.

In a polypeptide according to the invention said second domain preferably comprises a clostridial toxin heavy chain $H_{\text{\tiny N}}$ portion or a fragment or variant of a clostridial toxin heavy chain H_N portion. The fragment is optionally an N-terminal or C-terminal or internal fragment, so long as it retains the function of the $H_{\scriptscriptstyle N}$ Teachings of regions within the H_{N} responsible for its function are domain. provided for example in Biochemistry 1995, 34, pages 15175-15181 and Eur. J. Biochem, 1989, 185, pages 197-203. The variant has a different sequence from the H_N domain or fragment, though it too retains the function of the H_N domain. It is conveniently obtained by insertion, deletion and/or substitution of a H_N domain In embodiments of the invention, described below, it or fragment thereof. comprises (i) an N-terminal extension to a H_N domain or fragment, (ii) a C-terminal extension to a H_N domain or fragment, (iii) a modification to a H_N domain or fragment by alteration of at least one amino acid, or (iv) combinations of 2 or more of (i)-(iii). The clostridial toxin is preferably botulinum toxin or tetanus toxin.

The invention also provides a polypeptide comprising a clostridial neurotoxin light chain and a N-terminal fragment of a clostridial neurotoxin heavy chain, the fragment preferably comprising at least 423 of the N-terminal amino acids of the heavy chain of botulinum toxin type A, 417 of the N-terminal amino acids of the heavy chain of botulinum toxin type B or the equivalent number of N-terminal amino acids of the heavy chain of other types of clostridial toxin such that the fragment possesses an equivalent alignment of homologous amino acid residues.

These polypeptides of the invention are thus not composed of two or more polypeptides, linked for example by di-sulphide bridges into composite molecules. Instead, these polypeptides are single chains and are not active or their activity is

significantly reduced in an in vitro assay of neurotoxin endopeptidase activity.

Further, the polypeptides may be susceptible to be converted into a form exhibiting endopeptidase activity by the action of a proteolytic agent, such as trypsin. In this way it is possible to control the endopeptidase activity of the toxin light chain.

In a specific embodiment of the invention described in an example below, there is provided a polypeptide lacking a portion designated $H_{\rm C}$ of a clostridial toxin heavy chain. This portion, seen in the naturally produced toxin, is responsible for binding of toxin to cell surface receptors prior to internalisation of the toxin. This specific embodiment is therefore adapted so that it can not be converted into active toxin, for example by the action of a proteolytic enzyme. The invention thus also provides a polypeptide comprising a clostridial toxin light chain and a fragment of a clostridial toxin heavy chain, said fragment being not capable of binding to those cell surface receptors involved in the intoxicating action of clostridial toxin, and it is preferred that such a polypeptide lacks an intact portion designated $H_{\rm C}$ of a clostridial toxin heavy chain.

In further embodiments of the invention there are provided compositions containing a polypeptide comprising a clostridial toxin light chain and a portion designated H_N of a clostridial toxin heavy chain, and wherein the composition is free of clostridial toxin and free of any clostridial toxin precursor that may be converted into clostridial toxin by the action of a proteolytic enzyme. Examples of these compositions include those containing toxin light chain and H_N sequences of botulinum toxin types A, B, C₁, D, E, F and G.

The polypeptides of the invention are conveniently adapted to bind to, or include, a ligand for targeting to desired cells. The polypeptide optionally comprises a sequence that binds to, for example, an immunoglobulin. A suitable sequence is a tandem repeat synthetic IgG binding domain derived from domain B of Staphylococcal protein A. Choice of immunoglobulin specificity then determines the target for a polypeptide - immunoglobulin complex. Alternatively, the

polypeptide comprises a non-clostridial sequence that binds to a cell surface receptor, suitable sequences including insulin-like growth factor-1 (IGF-1) which binds to its specific receptor on particular cell types and the 14 amino acid residue sequence from the carboxy-terminus of cholera toxin A subunit which is able to bind the cholera toxin B subunit and thence to GM1 gangliosides. A polypeptide according to the invention thus, optionally, further comprises a third domain adapted for binding of the polypeptide to a cell.

In a second aspect the invention provides a fusion protein comprising a fusion of (a) a polypeptide of the invention as described above with (b) a second polypeptide adapted for binding to a chromatography matrix so as to enable purification of the fusion protein using said chromatography matrix. It is convenient for the second polypeptide to be adapted to bind to an affinity matrix, such as a glutathione Sepharose, enabling rapid separation and purification of the fusion protein from an impure source, such as a cell extract or supernatant.

One possible second purification polypeptide is glutathione-S-transferase (GST), and others will be apparent to a person of skill in the art, being chosen so as to enable purification on a chromatography column according to conventional techniques.

As noted above, by proteolytic treatment, for example using trypsin, of a polypeptide of the invention it is possible to induce endopeptidase activity in the treated polypeptide. A third aspect of the invention provides a composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the clostridial toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*. The activity of the derivative preferably approaches that of natural toxin, and is thus preferably at least 30% and most preferably at least 60% of natural toxin. The overall endopeptidase activity of the composition will, of course, also be determined by the amount of the derivative that is present.

While it is known to treat naturally produced clostridial toxin to remove the H_C domain, this treatment does not totally remove toxicity of the preparation, instead some residual toxin activity remains. Natural toxin treated in this way is therefore still not entirely safe. The composition of the invention, derived by treatment of a pure source of polypeptide advantageously is free of toxicity, and can conveniently be used as a positive control in a toxin assay, as a vaccine against clostridial toxin or for other purposes where it is essential that there is no residual toxicity in the composition.

The invention enables production of the polypeptides and fusion proteins of the invention by recombinant means.

A fourth aspect of the invention provides a nucleic acid encoding a polypeptide or a fusion protein according to any of the aspects of the invention described above.

In one embodiment of this aspect of the invention, a DNA sequence provided to code for the polypeptide or fusion protein is not derived from native clostridial sequences, but is an artificially derived sequence not preexisting in nature.

A specific DNA (SEQ ID NO: 1) described in more detail below encodes a polypeptide or a fusion protein comprising nucleotides encoding residues 1-871 of a botulinum toxin type A. Said polypeptide comprises the light chain domain and the first 423 amino acid residues of the amino terminal portion of a botulinum toxin type A heavy chain. This recombinant product is designated LH₄₂₃/A (SEQ ID NO: 2).

In a second embodiment of this aspect of the invention a DNA sequence which codes for the polypeptide or fusion protein is derived from native clostridial sequences but codes for a polypeptide or fusion protein not found in nature.

A specific DNA (SEQ ID NO: 19) described in more detail below encodes a polypeptide or a fusion protein and comprises nucleotides encoding residues 1-

1171 of a botulinum toxin type B. Said polypeptide comprises the light chain domain and the first 728 amino acid residues of the amino terminal protein of a botulinum type B heavy chain. This recombinant product is designated LH₇₂₈/B (SEQ ID NO: 20).

The invention thus also provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA according to the third aspect of the invention. The host cell is suitably not able to cleave a polypeptide or fusion protein of the invention so as to separate light and heavy toxin chains; for example, a non-clostridial host.

The invention further provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA encoding a fusion protein as described above, purifying the fusion protein by elution through a chromatography column adapted to retain the fusion protein, eluting through said chromatography column a ligand adapted to displace the fusion protein and recovering the fusion protein. Production of substantially pure fusion protein is thus made possible. Likewise, the fusion protein is readily cleaved to yield a polypeptide of the invention, again in substantially pure form, as the second polypeptide may conveniently be removed using the same type of chromatography column.

The LH_N/A derived from dichain native toxin requires extended digestion with trypsin to remove the C-terminal 1/2 of the heavy chain, the H_C domain. The loss of this domain effectively renders the toxin inactive *in vivo* by preventing its interaction with host target cells. There is, however, a residual toxic activity which may indicate a contaminating, trypsin insensitive, form of the whole type A neurotoxin.

In contrast, the recombinant preparations of the invention are the product of a discreet, defined gene coding sequence and can not be contaminated by full length toxin protein. Furthermore, the product as recovered from *E. coli*, and from other recombinant expression hosts, is an inactive single chain peptide or if expression

hosts produce a processed, active polypeptide it is not a toxin. Endopeptidase activity of LH_{423}/A , as assessed by the current *in vitro* peptide cleavage assay, is wholly dependent on activation of the recombinant molecule between residues 430 and 454 by trypsin. Other proteolytic enzymes that cleave between these two residues are generally also suitable for activation of the recombinant molecule. Trypsin cleaves the peptide bond C-terminal to Arginine or C-terminal to Lysine and is suitable as these residues are found in the 430-454 region and are exposed (see Fig. 12).

The recombinant polypeptides of the invention are potential therapeutic agents for targeting to cells expressing the relevant substrate but which are not implicated in effecting botulism. An example might be where secretion of neurotransmitter is inappropriate or undesirable or alternatively where a neuronal cell is hyperactive in terms of regulated secretion of substances other than neurotransmitter. In such an example the function of the H_c domain of the native toxin could be replaced by an alternative targeting sequence providing, for example, a cell receptor ligand and/or translocation domain.

One application of the recombinant polypeptides of the invention will be as a reagent component for synthesis of therapeutic molecules, such as disclosed in WO-A-94/21300. The recombinant product will also find application as a non-toxic standard for the assessment and development of *in vitro* assays for detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples, for example as disclosed in EP-A-0763131.

A further option is addition, to the C-terminal end of a polypeptide of the invention, of a peptide sequence which allows specific chemical conjugation to targeting ligands of both protein and non-protein origin.

In yet a further embodiment an alternative targeting ligand is added to the N-terminus of polypeptides of the invention. Recombinant LH_N derivatives have been designated that have specific protease cleavage sites engineered at the C-terminus

of the LC at the putative trypsin sensitive region and also at the extreme C-terminus of the complete protein product. These sites will enhance the activational specificity of the recombinant product such that the dichain species can only be activated by proteolytic cleavage of a more predictable nature than use of trypsin.

The LH_N enzymatically produced from native BoNT/A is an efficient immunogen and thus the recombinant form with its total divorce from any full length neurotoxin represents a vaccine component. The recombinant product may serve as a basal reagent for creating defined protein modifications in support of any of the above areas.

Recombinant constructs are assigned distinguishing names on the basis of their amino acid sequence length and their Light Chain (L-chain, L) and Heavy Chain (H-chain, H) content as these relate to translated DNA sequences in the public domain or specifically to SEQ ID NO: 2 and SEQ ID NO: 20. The 'LH' designation is followed by '/X' where 'X' denotes the corresponding clostridial toxin serotype or class, e.g. 'A' for botulinum neurotoxin type A or 'TeTx' for tetanus toxin. Sequence variants from that of the native toxin polypeptide are given in parenthesis in standard format, namely the residue position number prefixed by the residue of the native sequence and suffixed by the residue of the variant.

Subscript number prefixes indicate an amino-terminal (N-terminal) extension, or where negative a deletion, to the translated sequence. Similarly, subscript number suffixes indicate a carboxy terminal (C-terminal) extension or where negative numbers are used, a deletion. Specific sequence inserts such as protease cleavage sites are indicated using abbreviations, e.g. Factor Xa is abbreviated to FXa. L-chain C-terminal suffixes and H-chain N-terminal prefixes are separated by a / to indicate the predicted junction between the L and H-chains. Abbreviations for engineered ligand sequences are prefixed or suffixed to the clostridial L-chain or H-chain corresponding to their position in the translation product.

Following this nomenclature,

LH ₄₂₃ /A	= SEQ ID NO: 2, containing the entire L-chain and 423
	amino acids of the H-chain of botulinum neurotoxin type
	A ;

₂LH₄₂₃/A = a variant of this molecule, containing a two amino acid extension to the N-terminus of the L-chain;

 $_2L_{/2}H_{423}/A$ = a further variant in which the molecule contains a two amino acid extension on the N-terminus of both the L-chain and the H-chain;

²L_{FXa/2}H₄₂₃/A = a further variant containing a two amino acid extension to the N-terminus of the L-chain, and a Factor Xa cleavage sequence at the C-terminus of the L-chain which, after cleavage of the molecule with Factor Xa leaves a two amino acid N-terminal extension to the H-chain component; and

 $_2L_{FXa/2}H_{423}/A$ -IGF-1 = a variant of this molecule which has a further C-terminal extension to the H-chain, in this example the insulin-like growth factor 1 (IGF-1) sequence.

There now follows description of specific embodiments of the invention, illustrated by drawings in which:

Fig. 1 shows a schematic representation of the domain structure of botulinum neurotoxin type A (BoNT/A);

Fig. 2 shows a schematic representation of assembly of the gene for an embodiment of the invention designated LH₄₂₃/A;

- Fig. 3 is a graph comparing activity of native toxin, trypsin generated "native" LH_N/A and an embodiment of the invention designated ${}_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) in an *in vitro* peptide cleavage assay;
- Fig. 4 is a comparison of the first 33 amino acids in published sequences of native toxin and embodiments of the invention;
- Fig. 5 shows the transition region of an embodiment of the invention designated L/₄H₄₂₃/A illustrating insertion of four amino acids at the N-terminus of the H_N sequence; amino acids coded for by the *Eco* 47 III restriction endonuclease cleavage site are marked and the H_N sequence then begins ALN...;
- Fig. 6 shows the transition region of an embodiment of the invention designated $L_{FXa/3}H_{423}/A$ illustrating insertion of a Factor Xa cleavage site at the C-terminus of the L-chain, and three additional amino acids coded for at the N-terminus of the H-sequence; the N-terminal amino acid of the cleavage-activated H_N will be cysteine;
- Fig. 7 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated $L_{FXa/3}H_{423}/A$ -IGF-1, a fusion protein; the IGF-1 sequence begins at position G_{882} ;
- Fig. 8 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated $L_{FXa/3}H_{423}/A$ -CtxA14, a fusion protein; the C-terminal CtxA sequence begins at position Ω_{882} ;
- Fig.9 shows the C-terminal portion of the amino acid sequence of an

embodiment of the invention designated $L_{FXa/3}H_{423}/A-ZZ$, a fusion protein; the C-terminal ZZ sequence begins at position A_{890} immediately after a generase recognition site (underlined);

show schematic representations of manipulations of

Figs. 10 & 11 polypeptides of the invention; Fig. 10 shows LH₄₂₃/A with N-terminal addition of an affinity purification peptide (in this case GST) and C-terminal addition of an Ig binding domain; protease cleavage sites R1, R2 and R3 enable selective enzymatic separation of domains; Fig. 11 shows specific examples of protease cleavage sites R1, R2 and R3 and a C-terminal fusion peptide sequence;

Fig. 12 shows the trypsin sensitive activation region of a polypeptide of the invention;

shows Western blot analysis of recombinant LH₁₀₇/B expressed from *E.coli*; panel A was probed with anti-BoNT/B antiserum; Lane 1, molecular weight standards; lanes 2 & 3, native BoNT/B; lane 4, immunopurified LH₁₀₇/B; panel B was probed with anti-T7 peptide tag antiserum; lane 1, molecular weight standards; lanes 2 & 3, positive control *E.coli* T7 expression; lane 4 immunopurified LH₁₀₇/B.

The sequence listing that accompanies this application contains the following sequences:-

SEQ ID NO:

Sequence

1

DNA coding for LH₄₂₃/A

2	LH ₄₂₃ /A
:	
3	DNA coding for 23LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y), of which an
	N-terminal portion is shown in Fig. 4.
4	$_{23}LH_{423}/A (O_2E,N_{26}K,A_{27}Y)$
5	DNA coding for 2LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y), of which an N-
	terminal portion is shown in Fig.4
6	$_{2}LH_{423}/A (Q_{2}E, N_{26}K, A_{27}Y)$
7	DNA coding for native BoNT/A according to Binz et al
8	native BoNT/A according to Binz et al
9	DNA coding for L _{/4} H ₄₂₃ /A
10	L _{/4} H ₄₂₃ /A
11	DNA coding for L _{FXa} / ₃ H ₄₂₃ /A
12	L _{FXa} / ₃ H ₄₂₃ /A
13	DNA coding for L _{FXa} / ₃ H ₄₂₃ /A-IGF-1
14	L _{FXa} / ₃ H ₄₂₃ /A-IGF-1
15	DNA coding for L _{FXa} / ₃ H ₄₂₃ /A-CtxA14
16	$L_{FXa}/_3H_{423}/A$ -CtxA14
17	DNA coding for L _{FXa/3} H ₄₂₃ /A-ZZ
18	$L_{FXa/3}H_{423}/A-ZZ$
19	DNA coding for LH ₇₂₈ /B
20	LH ₇₂₈ /B
21	DNA coding for LH ₄₁₇ /B
22	LH ₄₁₇ /B
23	DNA coding for LH ₁₀₇ /B
24	LH ₁₀₇ /B
25	DNA coding for LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
26.	LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
27	DNA coding for LH ₄₁₇ /B wherein the first 274 bases are

28

modified to have an *E. coli* codon bias

DNA coding for LH₄₁₇/B wherein bases 691-1641 of the native BoNT/B sequence have been replaced by a degenerate DNA coding for amino acid residues 231-547 of the native BoNT/B polypeptide

Example 1

A 2616 base pair, double stranded gene sequence (SEQ ID NO: 1) has been assembled from a combination of synthetic, chromosomal and polymerase-chain-reaction generated DNA (Figure 2). The gene codes for a polypeptide of 871 amino acid residues corresponding to the entire light-chain (LC, 448 amino acids) and 423 residues of the amino terminus of the heavy-chain (H_c) of botulinum neurotoxin type A. This recombinant product is designated the LH₄₂₃/A fragment (SEQ ID NO: 2).

Construction of the recombinant product

The first 918 base pairs of the recombinant gene were synthesised by concatenation of short oligonucleotides to generate a coding sequence with an E. coli codon bias. Both DNA strands in this region were completely synthesised as short overlapping oligonucleotides which were phosphorylated, annealed and ligated to generate the full synthetic region ending with a unique Kpnl restriction site. The remainder of the LH_{423}/A coding sequence was PCR amplified from total chromosomal DNA from $Clostridium\ botulinum\$ and annealed to the synthetic portion of the gene.

The internal PCR amplified product sequences were then deleted and replaced with the native, fully sequenced, regions from clones of *C. botulinum* chromosomal origin to generate the final gene construct. The final composition is synthetic DNA (bases 1-913), polymerase amplified DNA (bases 914-1138 and 1976-2616) and the remainder is of *C. botulinum* chromosomal origin (bases 1139-1975). The

assembled gene was then fully sequenced and cloned into a variety of *E.coli* plasmid vectors for expression analysis.

Expression of the recombinant gene and recovery of protein product

The DNA is expressed in *E. coli* as a single nucleic acid transcript producing a soluble single chain polypeptide of 99,951 Daltons predicted molecular weight. The gene is currently expressed in *E. coli* as a fusion to the commercially available coding sequence of glutathione S-transferase (GST) of *Schistosoma japonicum* but any of an extensive range of recombinant gene expression vectors such as pEZZ18, pTrc99, pFLAG or the pMAL series may be equally effective as might expression in other prokaryotic or eukaryotic hosts such as the Gram positive bacilli, the yeast *P. pastoris* or in insect or mammalian cells under appropriate conditions.

Currently, E. coli harbouring the expression construct is grown in Luria-Bertani broth (L-broth pH 7.0, containing 10 g/l bacto-tryptone, 5 g/l bacto-yeast extract and 10 g/l sodium chloride) at 37° C until the cell density (biomass) has an optical absorbance of 0.4- 0.6 at 600 nm and the cells are in mid-logarithmic growth phase. Expression of the gene is then induced by addition isopropylthio- β -D-galactosidase (IPTG) to a final concentration of 0.5 mM. Recombinant gene expression is allowed to proceed for 90 min at a reduced temperature of 25°C. The cells are then harvested by centrifugation, are resuspended in a buffer solution containing 10 mM Na₂HPO₄, 0.5 M NaCl, 10 mM EGTA, 0.25% Tween, pH 7.0 and then frozen at -20°C. For extraction of the recombinant protein the cells are disrupted by sonication. The cell extract is then cleared of debris by centrifugation and the cleared supernatant fluid containing soluble recombinant fusion protein (GST- LH₄₂₃/A) is stored at -20°C pending purification. A proportion of recombinant material is not released by the sonication procedure and this probably reflects insolubility or inclusion body formation. Currently we do not extract this material for analysis but if desired this could be readily achieved using methods known to those skilled in the art.

The recombinant GST- LH_{423}/A is purified by adsorption onto a commercially prepared affinity matrix of glutathione Sepharose and subsequent elution with reduced glutathione. The GST affinity purification marker is then removed by proteolytic cleavage and reabsorption to glutathione Sepharose; recombinant LH_{423}/A is recovered in the non-adsorbed material.

Construct variants

A variant of the molecule, LH_{423}/A ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 26) has been produced in which three amino acid residues have been modified within the light chain of LH_{423}/A producing a polypeptide containing a light chain sequence different to that of the published amino acid sequence of the light chain of BoNT/A.

Two further variants of the gene sequence that have been expressed and the corresponding products purified are $_{23}LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 4) which has a 23 amino acid N-terminal extension as compared to the predicted native L-chain of BoNT/A and $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 6) which has a 2 amino acid N-terminal extension (Figure 4).

In yet another variant a gene has been produced which contains a Eco 47 III restriction site between nucleotides 1344 and 1345 of the gene sequence given in (SEQ ID NO: 1). This modification provides a restriction site at the position in the gene representing the interface of the heavy and light chains in native neurotoxin, and provides the capability to make insertions at this point using standard restriction enzyme methodologies known to those skilled in the art. It will also be obvious to those skilled in the art that any one of a number of restriction sites could be so employed, and that the Eco 47 III insertion simply exemplifies this approach. Similarly, it would be obvious for one skilled in the art that insertion of a restriction site in the manner described could be performed on any gene of the invention. The gene described, when expressed, codes for a polypeptide, $L_{/4}H_{423}/A$ (SEQ ID NO: 10), which contains an additional four amino acids between amino acids 448 and 449 of L_{H423}/A at a position equivalent to the amino terminus of the

heavy chain of native BoNT/A.

A variant of the gene has been expressed, L_{FXa/3}H₄₂₃/A (SEQ ID NO: 12), in which a specific proteolytic cleavage site was incorporated at the carboxy-terminal end of the light chain domain, specifically after residue 448 of L_{/4}H₄₂₃/A. The cleavage site incorporated was for Factor Xa protease and was coded for by modification of SEQ ID NO: 1. It will be apparent to one skilled in the art that a cleavage site for another specified protease could be similarly incorporated, and that any gene sequence coding for the required cleavage site could be employed. Modification of the gene sequence in this manner to code for a defined protease site could be performed on any gene of the invention.

Variants of $L_{FXa/3}H_{423}/A$ have been constructed in which a third domain is present at the carboxy-terminal end of the polypeptide which incorporates a specific binding activity into the polypeptide.

Specific examples described are:

- (1) $L_{FXa/3}H_{423}/A$ -IGF-1 (SEQ ID NO: 14), in which the carboxy-terminal domain has a sequence equivalent to that of insulin-like growth factor-1 (IGF-1) and is able to bind to the insulin-like growth factor receptor with high affinity;
- (2) $L_{FXa/3}H_{423}/A$ -CtxA14 (SEQ ID NO: 16), in which the carboxy-terminal domain has a sequence equivalent to that of the 14 amino acids from the carboxy-terminus of the A-subunit of cholera toxin (CtxA) and is thereby able to interact with the cholera toxin B-subunit pentamer; and
- (3) $L_{\rm FXa/3}H_{\rm 423}/A$ -ZZ (SEQ ID NO: 18), in which the carboxy-terminal domain is a tandem repeating synthetic IgG binding domain. This variant also exemplifies another modification applicable to the current invention, namely the inclusion in the gene of a sequence coding for a protease cleavage site located between the end of the clostridial heavy chain sequence and the sequence coding for the binding

ligand. Specifically in this example a sequence is inserted at nucleotides 2650 to 2666 coding for a generase cleavage site. Expression of this gene produces a polypeptide which has the desired protease sensitivity at the interface between the domain providing H_N function and the binding domain. Such a modification enables selective removal of the C-terminal binding domain by treatment of the polypeptide with the relevant protease.

It will be apparent that any one of a number of such binding domains could be incorporated into the polypeptide sequences of this invention and that the above examples are merely to exemplify the concept. Similarly, such binding domains can be incorporated into any of the polypeptide sequences that are the basis of this invention. Further, it should be noted that such binding domains could be incorporated at any appropriate location within the polypeptide molecules of the invention.

Further embodiments of the invention are thus illustrated by a DNA of the invention further comprising a desired restriction endonuclease site at a desired location and by a polypeptide of the invention further comprising a desired protease cleavage site at a desired location.

The restriction endonuclease site may be introduced so as to facilitate further manipulation of the DNA in manufacture of an expression vector for expressing a polypeptide of the invention; it may be introduced as a consequence of a previous step in manufacture of the DNA; it may be introduced by way of modification by insertion, substitution or deletion of a known sequence. The consequence of modification of the DNA may be that the amino acid sequence is unchanged, or may be that the amino acid sequence is changed, for example resulting in introduction of a desired protease cleavage site, either way the polypeptide retains its first and second domains having the properties required by the invention.

Figure 10 is a diagrammatic representation of an expression product exemplifying features described in this example. Specifically, it illustrates a single polypeptide

incorporating a domain equivalent to the light chain of botulinum neurotoxin type A and a domain equivalent to the H_N domain of the heavy chain of botulinum neurotoxin type A with a N-terminal extension providing an affinity purification domain, namely GST, and a C-terminal extension providing a ligand binding domain, namely an IgG binding domain. The domains of the polypeptide are spatially separated by specific protease cleavage sites enabling selective enzymatic separation of domains as exemplified in the Figure. This concept is more specifically depicted in Figure 11 where the various protease sensitivities are defined for the purpose of example.

Assay of product activity

The LC of botulinum neurotoxin type A exerts a zinc-dependent endopeptidase activity on the synaptic vesicle associated protein SNAP-25 which it cleaves in a specific manner at a single peptide bond. The $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 6) cleaves a synthetic SNAP-25 substrate *in vitro* under the same conditions as the native toxin (Figure 3). Thus, the modification of the polypeptide sequence of $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) relative to the native sequence and within the minimal functional LC domains does not prevent the functional activity of the LC domains.

This activity is dependent on proteolytic modification of the recombinant GST- $_2$ LH $_{423}$ /A (Q_2 E, N_{26} K, A_{27} Y) to convert the single chain polypeptide product to a disulphide linked dichain species. This is currently done using the proteolytic enzyme trypsin. The recombinant product (100-600 μ g/ml) is incubated at 37°C for 10-50 minutes with trypsin (10 μ g/ml) in a solution containing 140 mM NaCl, 2.7 mM KCl, 10 mM Na $_2$ HPO $_4$, 1.8 mM KH $_2$ PO $_4$, pH 7.3. The reaction is terminated by addition of a 100-fold molar excess of trypsin inhibitor. The activation by trypsin generates a disulphide linked dichain species as determined by polyacrylamide gel electrophoresis and immunoblotting analysis using polyclonal anti-botulinum neurotoxin type A antiserum.

₂LH₄₂₃/A is more stable in the presence of trypsin and more active in the in vitro

peptide cleavage assay than is 23LH423/A. Both variants, however, are fully functional in the *in vitro* peptide cleavage assay. This demonstrates that the recombinant molecule will tolerate N-terminal amino acid extensions and this may be expanded to other chemical or organic moieties as would be obvious to those skilled in the art.

Example 2

As a further exemplification of this invention a number of gene sequences have been assembled coding for polypeptides corresponding to the entire light-chain and varying numbers of residues from the amino terminal end of the heavy chain of botulinum neurotoxin type B. In this exemplification of the disclosure the gene sequences assembled were obtained from a combination of chromosomal and polymerase-chain-reaction generated DNA, and therefore have the nucleotide sequence of the equivalent regions of the natural genes, thus exemplifying the principle that the substance of this disclosure can be based upon natural as well as a synthetic gene sequences.

The gene sequences relating to this example were all assembled and expressed using methodologies as detailed in Sambrook J, Fritsch E F & Maniatis T (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Ford N, Nolan C, Ferguson M & Ockler M (eds), Cold Spring Harbor Laboratory Press, New York, and known to those skilled in the art.

A gene has been assembled coding for a polypeptide of 1171 amino acids corresponding to the entire light-chain (443 amino acids) and 728 residues from the amino terminus of the heavy chain of neurotoxin type B. Expression of this gene produces a polypeptide, LH₇₂₈/B (SEQ ID NO: 20), which lacks the specific neuronal binding activity of full length BoNT/B.

A gene has also been assembled coding for a variant polypeptide, LH_{417}/B (SEQ ID NO: 22), which possesses an amino acid sequence at its carboxy terminus

equivalent by amino acid homology to that at the carboxy-terminus of the heavy chain fragment in native $\ensuremath{\mathsf{LH}_{\mathsf{N}}/\mathsf{A}}$.

A gene has also been assembled coding for a variant polypeptide, LH_{107}/B (SEQ ID NO: 24), which expresses at its carboxy-terminus a short sequence from the amino terminus of the heavy chain of BoNT/B sufficient to maintain solubility of the expressed polypeptide.

Construct Variants

A variant of the coding sequence for the first 274 bases of the gene shown in SEQ ID NO: 21 has been produced which whilst being a non-native nucleotide sequence still codes for the native polypeptide.

Two double stranded, a 268 base pair and a 951 base pair, gene sequences have been created using an overlapping primer PCR strategy. The nucleotide bias of these sequences was designed to have an *E.coli* codon usage bias.

For the first sequence, six oligonucleotides representing the first (5') 268 nucleotides of the native sequence for botulinum toxin type B were synthesised. For the second sequence 23 oligonucleotides representing internal sequence nucleotides 691-1641 of the native sequence for botulinum toxin type B were synthesised. The oligonucleotides ranged from 57-73 nucleotides in length. Overlapping regions, 17-20 nucleotides, were designed to give melting temperatures in the range 52-56°C. In addition, terminal restriction endonuclease sites of the synthetic products were constructed to facilitate insertion of these products into the exact corresponding region of the native sequence. The 268 bp 5' synthetic sequence has been incorporated into the gene shown in SEQ ID NO: 21 in place of the original first 268 bases (and is shown in SEQ ID NO: 27). Similarly the sequence could be inserted into other genes of the examples.

Another variant sequence equivalent to nucleotides 691 to 1641 of SEQ ID NO: 21

, and employing non-native codon usage whilst coding for a native polypeptide sequence, has been constructed using the internal synthetic sequence. This sequence (SEQ ID NO: 28) can be incorporated, alone or in combination with other variant sequences, in place of the equivalent coding sequence in any of the genes of the example.

Example 3

An exemplification of the utility of this invention is as a non-toxic and effective immunogen. The non-toxic nature of the recombinant, single chain material was demonstrated by intraperitoneal administration in mice of GST-2LH423/A. The polypeptide was prepared and purified as described above. The amount of immunoreactive material in the final preparation was determined by enzyme linked immunosorbent assay (ELISA) using a monoclonal antibody (BA11) reactive against a conformation dependent epitope on the native LH_N/A. The recombinant material was serially diluted in phosphate buffered saline (PBS; NaCl 8 g/l, KCl 0.2 g/l, Na₂HPO₄ 1.15 g/l, KH₂PO₄ 0.2 g/l, pH 7.4) and 0.5 ml volumes injected into 3 groups of 4 mice such that each group of mice received 10, 5 and 1 micrograms of material respectively. Mice were observed for 4 days and no deaths were seen.

For immunisation, 20 μ g of GST-₂LH₄₂₃/A in a 1.0 ml volume of water-in-oil emulsion (1:1 vol:vol) using Freund's complete (primary injections only) or Freund's incomplete adjuvant was administered into guinea pigs via two sub-cutaneous dorsal injections. Three injections at 10 day intervals were given (day 1, day 10 and day 20) and antiserum collected on day 30. The antisera were shown by ELISA to be immunoreactive against native botulinum neurotoxin type A and to its derivative LH_N/A. Antisera which were botulinum neurotoxin reactive at a dilution of 1:2000 were used for evaluation of neutralising efficacy in mice. For neutralisation assays 0.1 ml of antiserum was diluted into 2.5 ml of gelatine phosphate buffer (GPB; Na₂HPO₄ anhydrous 10 g/I, gelatin (Difco) 2 g/I, pH 6.5-6.6) containing a dilution range from 0.5 μ g (5X10-6 g) to 5 picograms (5X10-12 g). Aliquots of 0.5 ml were injected into mice intraperitoneally and deaths recorded

over a 4 day period. The results are shown in Table 1 and Table 2. It can clearly be seen that 0.5 ml of 1:40 diluted anti- $GST_{-2}LH_{423}/A$ antiserum can protect mice against intraperitoneal challenge with botulinum neurotoxin in the range 5 pg - 50 ng (1 - 10,000 mouse LD50; 1 mouse LD50 = 5 pg).

TABLE 1. Neutralisation of botulinum neurotoxin in mice by guinea pig anti-GST-2LH423/A antiserum.

		=	otomiani tox				
Survivors On Day	0.5µg	0.005μg	0.0005µg	0.5ng	0.005ng	5pg	Control
1	0	4	4	4	4	4	4
2	-	4	4	4	4	4	4
3	-	4	4	4	4	4	4
4	-	4.	4	4	4	4	4

Botulinum Toxin/mouse

TABLE 2. Neutralisation of botulinum neurotoxin in mice by non-immune guinea pig antiserum.

Botulinum Toxin/mouse													
Survivors On Day	0.5µg	0.005µg	0.0005µg	0.5ng	0.005ng	5pg	Control (no toxin)						
1	0	0	0	0	0	2	4						
2	-	-	•	-	-	0	4						
3 .	-	-	•	-	-	-	. 4						
4	-	-	•	-	-		4						

Example 4

Expression of recombinant LH₁₀₇/B in E. coli.

As an exemplification of the expression of a nucleic acid coding for a LH_N of a clostridial neurotoxin of a serotype other than botulinum neurotoxin type A, the nucleic acid sequence (SEQ ID NO: 23) coding for the polypeptide LH_{107}/B (SEQ ID

NO: 24) was inserted into the commercially available plasmid pET28a (Novogen, Madison, WI, USA). The nucleic acid was expressed in $E.\ coli$ BL21 (DE3) (New England BioLabs, Beverley, MA, USA) as a fusion protein with a N-terminal T7 fusion peptide, under IPTG induction at 1 mM for 90 minutes at 37°C. Cultures were harvested and recombinant protein extracted as described previously for LH_{423}/A .

Recombinant protein was recovered and purified from bacterial paste lysates by immunoaffinity adsorption to an immobilised anti-T7 peptide monoclonal antibody using a T7 tag purification kit (New England bioLabs, Beverley, MA, USA). Purified recombinant protein was analysed by gradient (4-20%) denaturing SDS-polyacrylamide gel electrophoresis (Novex, San Diego, CA, USA) and western blotting using polyclonal anti-botulinum neurotoxin type antiserum or anti-T7 antiserum. Western blotting reagents were from Novex, immunostained proteins were visualised using the Enhanced Chemi-Luminescence system (ECL) from Amersham. The expression of an anti-T7 antibody and anti-botulinum neurotoxin type B antiserum reactive recombinant product is demonstrated in Figure 13.

The recombinant product was soluble and retained that part of the light chain responsible for endopeptidase activity.

The invention thus provides recombinant polypeptides useful inter alia as immunogens, enzyme standards and components for synthesis of molecules as described in WO-A-94/21300.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT:
 - (A) NAME: MICROBIOLOGICAL RESEARCH AUTHORITY
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) NAME: THE SPEYWOOD LABORATORY LIMITED
 - (B) STREET: 14 Kensington Square
 - (C) CITY: London (E) COUNTRY: UK

 - (F) POSTAL CODE (ZIP): W8 5HH
 - (A) NAME: FOSTER; Keith Alan
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) NAME: QUINN; Conrad Padraig
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) NAME: SHONE; Clifford Charles
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
- (ii) TITLE OF INVENTION: Recombinant Toxin Fragments
- (iii) NUMBER OF SEQUENCES: 28
- (iv) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
- (2) INFORMATION FOR SEQ ID NO: 1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2616 base pairs

 - (B) TYPE: nucleic acid(C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

							-									
ATO Met		TT(C GT(e Val	G AA(L Asr	Lys	Gln	TTO Phe	AAC Asr	TA: 1 Ty:	r Lys	G GAG S Asp	CC1	GTZ Val	A AA L Asi	C GGT n Gly	4 6
GTT Val	GAC Asp	ATT	GCC Ala 20	TAT	Ile	Lys	ATT Ile	CCA Pro 25) Asr	GCC Ala	GGC Gly	CAG Gln	ATO Met	Glı	G CCG	96
GTG Val	AAG Lys	GCT Ala 35		AAG Lys	ATT Ile	CAT	AAC Asn 40	r	ATC	TGG Trp	GTT Val	ATT Ile 45	Pro	GAZ Glu	CGC Arg	144
GAT Asp	ACA Thr 50	TTT Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	Pro	CCG Pro	CCC Pro	GAA Glu	192
GCA Ala 65	27.5	CAG Gln	GTG Val	CCA Pro	GTT Val 70	TCA Ser	TAC Tyr	TAC	GAT Asp	TCA Ser 75	ACC Thr	TAT Tyr	CTG Leu	AGC Ser	ACA Thr 80	240
пэр	ASII	GIU	цуѕ	85	ASII	ıyr	Leu	Lys	90 90	Val	Thr	Lys	Leu	Phe 95	GAG Glu	288
CGT Arg	ATT Ile	TAT Tyr	TCC Ser 100	ACT Thr	GAC Asp	CTG Leu	GGC Gly	CGT Arg 105	ATG Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	ATC Ile	GTC Val	336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys	384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	Thr	GIn	TAC Tyr 185	Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	Val	ACC Thr 220	CTG(Leu)	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GIY	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	Gly	ATT (Ile / 235	GCG Ala	ATT A	AAC Asn	CCG Pro	AAC Asn 240	720

CGC GTG 1 Arg Val I	TTC AAG GTT AA Phe Lys Val As 245	C ACC AAC GC n Thr Asn Al	C TAC TAC GAG A a Tyr Tyr Glu ! 250	ATG AGT GGT TTA Met Ser Gly Leu 255	768
GAA GTA A Glu Val S	AGC TTC GAG GA Ser Phe Glu Gl 260	A CTG CGC ACC u Leu Arg Th 26!	r Phe Gly Gly F	CAT GAT GCG AAG His Asp Ala Lys 270	816
Pne lle A	AC AGC TTG CA sp Ser Leu Gl 75	G GAG AAC GAG n Glu Asn Glu 280	1 Phe Arg Leu I	TAC TAC TAC AAC Yr Tyr Tyr Asn	864
AAG TTT A Lys Phe L 290	AA GAT ATT GC: ys Asp Ile Al:	A AGT ACA CTO A Ser Thr Leu 295	G AAC AAG GCT A n Asn Lys Ala L 300	AG TCC ATT GTG ys Ser Ile Val	912
GGT ACC AGGLY Thr TI	CT GCT TCA TTA hr Ala Ser Let 310	l Gin Tyr Met	AAA AAT GTT T Lys Asn Val P 315	TT AAA GAG AAA he Lys Glu Lys 320	960
TAT CTC CT	TA TCT GAA GAT eu Ser Glu Asp 325	ACA TCT GGA Thr Ser Gly	AAA TTT TCG G Lys Phe Ser V 330	TA GAT AAA TTA al Asp Lys Leu 335	1008
AAA TTT GA Lys Phe As	AT AAG TTA TAC Sp Lys Leu Tyr 340	-AAA ATG TTA Lys Met Leu 345	ACA GAG ATT TA	AC ACA GAG GAT yr Thr Glu Asp 350	1056
AAT TTT GT Asn Phe Va 35	ii ras bue bue	AAA GTA CTT Lys Val Leu 360	AAC AGA AAA AG Asn Arg Lys Th	nr Tyr Leu Asn	1104
TTT GAT AA Phe Asp Ly 370	A GCC GTA TTT 'S Ala Val Phe	AAG ATA AAT Lys Ile Asn 375	ATA GTA CCT AF Ile Val Pro Ly 380	AG GTA AAT TAC	1152
ACA ATA TA Thr Ile Ty 385	T GAT GGA TTT r Asp Gly Phe 390	AAT TTA AGA Asn Leu Arg	AAT ACA AAT TT Asn Thr Asn Le 395	A GCA GCA AAC u Ala Ala Asn 400	1200
TTT AAT GG Phe Asn Gl	T CAA AAT ACA y Gln Asn Thr 405	GAA ATT AAT Glu Ile Asn	AAT ATG AAT TT Asn Met Asn Ph 410	T ACT AAA CTA e Thr Lys Leu 415	1248
AAA AAT TT Lys Asn Ph	e Inr Gly Leu	TTT GAA TTT Phe Glu Phe 425	TAT AAG TTG CT Tyr Lys Leu Le	A TGT GTA AGA u Cys Val Arg 430	1296
GGG ATA ATA Gly Ile Ile 439	e Thr Ser Lys	ACT AAA TCA Thr Lys Ser 440	TTA GAT AAA GG Leu Asp Lys Gl 449	y Tyr Asn Lys	1344
GCA TTA AAT Ala Leu Ass 450	GAT TTA TGT Asp Leu Cys	ATC AAA GTT Ile Lys Val 455	AAT AAT TGG GAG Asn Asn Trp Asp 460	C TTG TTT TTT D Leu Phe Phe	1392
AGT CCT TCA Ser Pro Ser 465	A GAA GAT AAT Glu Asp Asn 470	TTT ACT AAT Phe Thr Asn	GAT CTA AAT AAA Asp Leu Asn Lys 475	A GGA GAA GAA 5 Gly Glu Glu 480	1440
ATT ACA TCT Ile Thr Ser	GAT ACT AAT Asp Thr Asn 485	Ile Glu Ala i	GCA GAA GAA AAT Ala Glu Glu Asn 490	TATT AGT TTA Ile Ser Leu 495	1488
GAT TTA ATA Asp Leu Ile	CAA CAA TAT Gln Gln Tyr 500	TAT TTA ACC T Tyr Leu Thr I 505	PTT AAT TTT GAT Phe Asn Phe Asp	AAT GAA CCT Asn Glu Pro 510	1536

-	32	-
---	----	---

						AAT Asn										1584
						GAA Glu 535									GAG Glu	1632
TTA Leu 545	GAT Asp	AAA Lys	TAT Tyr	ACT Thr	ATG Met 550	TTC Phe	CAT His	TAT Tyr	CTT Leu	CGT Arg 555	GCT Ala	CAA Gln	GAA Glu	TTT Phe	GAA Glu 560	1680
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	ATT Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT	GTT Val	AAC Asn	GAA Glu	GCA Ala 575	TTA Leu	1728
						TAT Tyr									AAG Lys	1776
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT Ala	ACG Thr	GAG Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	Trp	GTA Val	GAA Glu	1824
CAA Gln	TTA Leu 610	GTA Val	TAT Tyr	GAT Asp	TTT Phe	ACC Thr 615	Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr	1872
						ACT Thr										1920
TTA Leu	AAT Asn	ATA Ile	GGT Gly	AAT Asn 645	ATG Met	TTA Leu	TAT Tyr	AAA Lys	GAT Asp 650	GAT Asp	TTT Phe	GTA Val	GGT Gly	GCT Ala 655	TTA Leu	1968
ATA Ile	TTT Phe	TCA Ser	GGA Gly 660	GCT Ala	GTT Val	ATT Ile	CTG Leu	TTA Leu 665	GAA Glu	TTT Phe	ATA Ile	CCA Pro	GAG Glu 670	ATT Ile	GCA Ala	2016
ATA Ile	CCT Pro	GTA Val 675	TTA Leu	GGT Gly	ACT Thr	TTT Phe	GCA Ala 680	CTT Leu	GTA Val	TCA Ser	TAT Tyr	ATT Ile 685	GCG Ala	AAT Asn	AAG Lys	2064
GTT Val	CTA Leu 690	ACC Thr	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	AAA Lys	AGA Arg	AAT Asn	GAA Glu	2112
AAA Lys 705	TGG Trp	GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	AAA Lys	TAT	ATA Ile	GTA Val	ACA Thr 715	AAT Asn	TGG Trp	TTA Leu	GCA Ala	AAG Lys 720	2160
GTT Val	AAT Asn	ACA Thr	CAG Gln	ATT Ile 725	GAT Asp	CTA Leu	ATA Ile	AGA Arg	AAA Lys 730	AAA Lys	ATG Met	AAA Lys	GAA Glu	GCT Ala 735	TTA Leu	2208
GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	AAT Asn	2256
CAA Gln	TAT Tyr	ACT Thr 755	Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT Ile	GAT Asp	GAT Asp	2304
TTA Leu	AGT Ser 770	Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	Met	ATT Ile	AAT Asn	ATA Ile	2352

AAT Asn 785	гÀг	TTT Phe	TTG Leu	AAT Asn	CAA Gln 790	TGC Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800	2	2,400
ATC Ile	CCT Pro	TAT	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GAA Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys	2	448
GAT Asp	GCA Ala	TTA Leu	TTA Leu 820	AAG Lys	TAT Tyr	ATA Ile	TAT Tyr	GAT Asp 825	AAT Asn	AGA Arg	GGA Gly	ACT Thr	TTA Leu 830	ATT Ile	GGT Gly	2	496
CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp	2	544
Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	TCC Ser	AAA Lys 855	TAC Tyr	GTA Val	GAT Asp	AAT Asn	CAA Gln 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser	2	592
			GAA Glu				TAA *									26	616

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 872 amino acids
 - (B) TYPE: amino acid(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 170

- 34 -

Arg	Asn	Gly	Tyr 180	Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	Phe	Ser	Pro 190	Asp	Phe
Thr	Phe	Gly 195	Phe	Glu	Glu	Ser	Leu 200	Glu	Val	Asp		Asn .205	Pro	Leu	Leu
Gly	Ala 210	Gly	Lys	Phe	Ala	Thr 215	Asp	Pro	Ala	Val	Thr 220	Leu	Ala	His	Glu
Leu 225	Ile	His	Ala	Gly	His 230	Arg	Leu	Tyr	Gly	Ile 235	Ala	Ile	Asn	Pro	Asn 240
Arg	Val	Phe	Lys	Val 245	Asn	Thr	Asn	Ala	Tyr 250	Tyr	Glu	Met	Ser	Gly 255	Leu
Glu	Val	Ser	Phe 260	Glu	Glu	Leu	Arg	Thr 265	Phe	Gly	Gly	His	Asp 270	Ala	Lys
Phe	Ile	Asp 275	Ser	Leu	Gln	Glu	Asn 280	Glu	Phe	Arg	Leu	Tyr 285	Tyr	Tyr	Asn
Lys	Phe 290	Lys	Asp	Ile	Ala	Ser 295		Leu	Asn	Lys	Ala 300	Lys	Ser	Ile	Val
Gly 305	Thr	Thr	Ala	Ser	Leu 310	Gln	Tyr	Met	Lys	Asn 315	Val	Phe	Lys	Glu	Lys 320
Tyr	Leu	Leu	Ser	Glu 325	Asp	Thr	Ser	Gly	Lys 330	Phe	Ser	Val	Asp	Lys 335	Leu
Lys	Phe	Asp	Lys 340	Leu	Tyr	Lys	Met	Leu 345	Thr	Glu	Ile	Tyr	Thr 350	Glu	Asp
Asn	Phe	Val 355	Lys	Phe	Phe	Lys	Val 360	Leu	Asn	Arg	Lys	Thr 365	Tyr	Leu	Asn
Phe	Asp 370	Lys	Ala	Val	Phe	Lys 375	Ile	Asn	Ile	Val	Pro 380	Lys	Val	Asn	Tyr
Thr 385	Ile	Tyr	Asp	Gly	Phe 390	Asn	Leu	Arg	Asn	Thr 395	Asn	Leu	Ala	Ala	Asn 400
Phe	Asn	Gly	Gln	Asn 405	Thr	Glu	Ile	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415	Leu
Lys	Asn	Phe	Thr 420	Gly	Leu	Phe	Glu	Phe 425	Tyr	Lys	Leu	Leu	Cys 430	Val	Arg
Gly	Ile	Ile 435	Thr	Ser	Lys	Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	Lys
Ala	Leu 450	Asn	Asp	Leu	Cys	Ile 455	Lys	Val	Asn	Asn	Trp 460	Asp	Leu	Phe	Phe
Ser 465	Pro	Ser	Glu	Asp	Asn 470	Phe	Thr	Asn	Asp	Leu 475	Asn	Lys	Gly	Glu	Glu 480
Ile	Thr	Ser	Asp	Thr 485	Asn	Ile	Glu	Ala	Ala 490	Glu	Glu	Asn	Ile	Ser 495	Leu
Asp	Leu	Ile	Gln 500	Gln	Tyr	Tyr	Leu	Thr 505	Phe	Asn	Phe	Asp	Asn 510	Glu	Pro
Glu	Asn	Ile 515		Ile	Glu	Asn	Leu 520	Ser	Ser	Asp	Ile	Ile 525	Gly	Gln	Leu

Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 530 540

Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 550 555 560

His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 575

Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys
580 585 590

Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu
595 600 605

Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 620

Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635 640

Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 655

Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670

Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 685

Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu 690 695 700

Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys
705 710 715 720

Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 725 730 735

Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn 740 745 750

Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp 755 760 765

Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 770 775 780

Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met 785 790 795 800

Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phē Asp Ala Ser Leu Lys 805 810 815

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 820 825 830

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 850 855

Thr Phe Thr Glu Tyr Ile Lys 865 870

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2685 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double

 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..2685
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGA Gly 1	TCC Ser	CCA Pro	GGA Gly	ATT Ile 5	CAT His	ATG Met	ACG Thr	TCG Ser	ACG Thr 10	CGT Arg	CTG Leu	CAG Gln	AAG Lys	CTT Leu 15	CTA Leu	48
GAA Glu	TTC Phe	GAG Glu	CTC Leu 20	CCG Pro	GGT Gly	ACC Thr	ATG Met	GAG Glu 25	TTC Phe	GTG Val	AAC Asn	AAG Lys	CAG Gln 30	TTC Phe	AAC Asn	96
TAT Tyr	AAG Lys	GAC Asp 35	CCT Pro	GTA Val	AAC Asn	GGT Gly	GTT Val 40	GAC Asp	ATT Ile	GCC Ala	TAC Tyr	ATC Ile 45	AAA Lys	ATT Ile	CCA Pro	144
		GGC Gly														192
		GTT Val														240
		AAC Asn														288
		ACC Thr														336
		ACC Thr 115														384
		CTG Leu														432
		GAC Asp					_									480
		CCA Pro														528
		CCC Pro														576
		GTG Val 195														624

ATT Ile	CGT Arg	g Phe	C AGO	CCA Pro	A GAC	TTC Phe 215	Thr	TT(GGT Gly	TTC Phe	GAG Glu 220	ı Gl	G AG	C CT	G GAG u Glu	672
GT1 Val 225	. Asp	ACC Thi	AAC Asn	CCG Pro	CTG Leu 230	Leu	GGT Gly	GCA Ala	GGC Gly	Lys 235	Phe	GC/ Ala	A AC	r GA'	CCA P Pro 240	720
GCG Ala	GTG Val	ACC Thr	CTG Leu	GCA Ala 245	His	GAG Glu	CTG Leu	ATC	CAC His 250	Ala	GG1 Gly	CAT His	CG1 Arg	CTC Let 255	TAT Tyr	768
GGC Gly	ATT	GCG Ala	Ile 260	AAC Asn	CCG Pro	AAC Asn	CGC Arg	GTG Val 265	TTC Phe	AAG Lys	GTT Val	AAC Asr	ACC Thr 270	Ası	GCC Ala	816
TAC Tyr	TAC Tyr	GAG Glu 275	Met	AGT Ser	GGT Gly	TTA Leu	GAA Glu 280	GTA Val	AGC Ser	TTC Phe	GAG Glu	GAA Glu 285	Leu	CGC Arg	ACG Thr	864
TTC Phe	GGT Gly 290	GGC Gly	CAT	GAT Asp	GCG Ala	AAG Lys 295	TTT Phe	ATC Ile	GAC Asp	AGC Ser	TTG Leu 300	CAG Gln	GAG Glu	AAC Asn	GAG Glu	912
TTC Phe 305	CGT Arg	CTG Leu	TAC Tyr	TAC Tyr	TAC Tyr 310	AAC Asn	AAG Lys	TTT Phe	AAA Lys	GAT Asp 315	ATT Ile	GCA Ala	AGT Ser	ACA Thr	CTG Leu 320	960
AAC Asn	AAG Lys	GCT Ala	AAG Lys	TCC Ser 325	ATT Ile	GTG Val	GGT Gly	ACC Thr	ACT Thr 330	GCT Ala	TCA Ser	TTA Leu	CAG Gln	TAT Tyr 335	ATG Met	1008
AAA Lys	AAT Asn	GTT Val	TTT Phe 340	AAA Lys	GAG Glu	AAA Lys	TAT Tyr	CTC Leu 345	CTA Leu	TCT Ser	GAA Glu	GAT Asp	ACA Thr 350	TCT Ser	GGA Gly	1056
AAA Lys	TTT Phe	TCG Ser 355	GTA Val	Așp GAT	AAA Lys	TTA Leu	AAA Lys 360	TTT Phe	GAT Asp	AA G Lys	TTA Leu	TAC Tyr 365	AAA Lys	ATG Met	TTA Leu	1104
ACA Thr	GAG Glu 370	ATT Ile	TAC Tyr	ACA Thr	GAG Glu	GAT Asp 375	AAT Asn	TTT Phe	GTT Val	AAG Lys	TTT Phe 380	TTT Phe	AAA Lys	GTA Val	CTT Leu	1152
Asn	Arg	Lys	ACA Thr	Tyr	Leu	TAA neA	Phe	Asp	AAA Lys	Ala	GTA Val	TTT Phe	AAG Lys	ATA Ile	AAT Asn 400	1200
ATA Ile	GTA Val	CCT Pro	AAG Lys	GTA Val 405	AAT Asn	TAC Tyr	ACA Thr	ATA Ile	TAT Tyr 410	GAT Asp	GGA Gly	TTT Phe	AAT Asn	TTA Leu 415	AGA Arg	1248
AAT Asn	ACA Thr	AAT Asn	TTA Leu 420	GCA Ala	GCA Ala	AAC ' Asn	Phe .	AAT Asn 425	GGT Gly	CAA Gln	AAT Asn	ACA Thr	GAA Glu 430	ATT Ile	AAT Asn	1296
AAT Asn	ATG Met	AAT Asn 435	TTT Phe	ACT Thr	AAA Lys	CTA .	AAA Lys 440	AAT Asn	TTT Phe	ACT Thr	GGA Gly	TTG Leu 445	TTT Phe	GAA Glu	TTT Phe	1344
Tyr	AAG Lys 450	TTG Leu	CTA Leu	TGT Cys	Val .	AGA (Arg (455	GGG :	ATA . Ile	ATA Ile	Thr	TCT Ser 460	A AA Lys	ACT Thr	AAA Lys	TCA Ser	1392
TTA Leu 465	GAT Asp	AAA Lys	GGA Gly	Tyr .	AAT Asn 470	AAG (Lys)	GCA 1	ITA . Leu .	Asn .	GAT ' Asp 1 475	TTA ' Leu '	TGT Cys	ATC Ile	AAA Lys	GTT Val 480	1440

AA1 Asr	TAA 1 naA n	TGG Trp	GAC Asp	TTG Leu 485	. Pne	TTT Phe	AGT Ser	Pro	TCA Ser 490	Glu	GAT LAST	AAT Asn	TTT Phe	TACT	AAT Asn	1488	8
GA1 Asp	CTA Leu	AAT Asn	AAA Lys 500	GrA	GAA Glu	GAA Glu	ATT Ile	ACA Thr 505	TCT Ser	GAT Asp	ACT Thr	AAT Asn	ATA Ile 510	: Glu	GÇA Ala	1536	5
Ата	~	515	Asn	ile	ser	Leu	Asp 520	Leu	Ile	Gln	Gln	Tyr 525	Tyr	Leu	ACC Thr	1584	1
Pne	530	Pne	Asp	Asn	GIu	Pro 535	Glu	Asn	Ile	Ser	11e 540	Glu	Asn	Leu	TCA Ser	1632	!
545	Asp	116	116	GIY	550	Leu	Glu	Leu	Met	Pro 555	Asn	Ile	Glu	Arg	TTT Phe 560	1680)
Pro	Asn	GGA Gly	Lys	565	Tyr	Glu	Leu	Asp	Lys 570	Tyr	Thr	Met	Phe	His 575	Tyr	1728	
CTT Leu	CGT Arg	GCT Ala	CAA Gln 580	GAA Glu	TTT Phe	GAA Glu	CAT His	GGT Gly 585	AAA Lys	TCT Ser	AGG Arg	ATT Ile	GCT Ala 590	TTA Leu	ACA Thr	1776	
AAT Asn	TCT Ser	GTT Val 595	AAC Asn	GAA Glu	GCA Ala	TTA Leu	TTA Leu 600	Asn	CCT Pro	AGT Ser	CGT Arg	GTT Val 605	TAT Tyr	ACA Thr	TTT Phe	1824	
TTT Phe	TCT Ser 610	TCA Ser	GAC Asp	TAT Tyr	GTA Val	AAG Lys 615	AAA Lys	GTT Val	AAT Asn	AAA Lys	GCT Ala 620	ACG Thr	GAG Glu	GCA Ala	GCT Ala	1872	
ATG Met 625	TTT Phe	TTA Leu	GGC Gly	TGG Trp	GTA Val 630	GAA Glu	CAA Gln	TTA Leu	GTA Val	TAT Tyr 635	GAT Asp	TTT Phe	ACC Thr	GAT Asp	GAA Glu 640	1920	•
ACT Thr	AGC Ser	GAA Glu	GTA Val	AGT Ser 645	ACT Thr	ACG Thr	GAT Asp	AAA Lys	ATT Ile 650	GCG Ala	GAT Asp	ATA Ile	ACT Thr	ATA Ile 655	ATT	1968	
ATT Ile	CCA Pro	TAT	ATA Ile 660	GGA Gly	CCT Pro	GCT Ala	Leu	AAT Asn 665	ATA Ile	GGT Gly	AAT Asn	ATG Met	TTA Leu 670	TAT Tyr	AAA Lys	2016	
GAT Asp	GAT Asp	TTT Phe 675	GTA Val	GGT Gly	GCT Ala	TTA Leu	ATA Ile 680	TTT Phe	TCA Ser	GGA Gly	GCT Ala	GTT Val 685	ATT Ile	CTG Leu	TTA Leu	2064	
GAA Glu	TTT Phe 690	ATA Ile	CCA Pro	GAG Glu	ATT Ile	GCA Ala 695	ATA Ile	CCT Pro	GTA Val	TTA Leu	GGT Gly 700	ACT Thr	TTT Phe	GCA Ala	CTT Leu	2112	
GTA Val 705	TCA Ser	TAT Tyr	ATT Ile	GCG Ala	AAT Asn 710	AAG Lys	GTT Val	CTA Leu	ACC Thr	GTT Val 715	CAA Gln	ACA Thr	ATA Ile	GAT Asp	AAT Asn 720	2160	
GCT Ala	TTA Leu	AGT Ser	AAA Lys	AGA Arg 725	AAT Asn	GAA Glu	AAA Lys	TGG Trp	GAT Asp 730	GAG Glu	GTC Val	TAT Tyr	AAA Lys	TAT Tyr 735	ATA Ile	2208	
GTA Val	ACA Thr	AAT Asn	TGG Trp 740	TTA Leu	GCA Ala	AAG Lys	Val	AAT Asn 745	ACA Thr	CAG Gln	ATT Ile	GAT Asp	CTA Leu 750	ATA Ile	AGA Arg	2256	

· Ly :	, Dys	755	Lys	GIU	MIG	Leu	760	ASI	ı GII	ı Ala	Gli	1 Ala 765	Thr	Ly	G GCT S Ala		2304
ATA Ile	ATA Ile 770	. ASII	TAT	CAG	TAT	AAT Asn 775	CAA Gln	TAT	ACT Thr	GAG Glu	GAA Glu 780	ı Glu	AAA Lys	AA7 Asr	AAT Asn		2352
ATT Ile 785	ASII	TTT Phe	AAT Asn	ATT	GAT Asp 790	GAT Asp	TTA Leu	AGT Ser	TCG Ser	AAA Lys 795	CTT Leu	AAT Asn	GAG Glu	TCT	ATA Ile 800	;	2400 [.]
ASII	rys	Ala	Mec	805	ASN	11e	Asn	Lys	Phe 810	Leu	Asn	Gln	Cys	Ser 815		2	2448
TCA Ser	TAT	TTA Leu	ATG Met 820	AAT Asn	TCT	ATG Met	ATC Ile	CCT Pro 825	TAT Tyr	GGT Gly	GTT Val	AAA Lys	CGG Arg 830	TTA Leu	GAA Glu	2	496
GAT Asp	TTT Phe	GAT Asp 835	GCT Ala	AGT Ser	CTT Leu	AAA Lys	GAT Asp 840	GCA Ala	TTA Leu	TTA Leu	AAG Lys	TAT Tyr 845	ATA Ile	TAT Tyr	GAT Asp	2	544
AAT Asn	AGA Arg 850	GGA Gly	ACT Thr	TTA Leu	TIE	GGT Gly 855	CAA Gln	GTA Val	GAT Asp	AGA Arg	TTA Leu 860	AAA Lys	GAT Asp	AAA Lys	GTT Val	2	592
AAT Asn 865	AAT Asn	ACA Thr	CTT Leu	Ser	ACA Thr 870	GAT Asp	ATA Ile	CCT Pro	Phe	CAG Gln 875	CTT Leu	TCC Ser	AAA Lys	TAC Tyr	GTA Val 880	2	640
GAT Asp	AAT Asn	CAA Gln	AGA (TTA Leu 885	TTA Leu	TCT / Ser '	ACA Thr	Phe	ACT Thr 890	GAA Glu	TAT Tyr	ATT Ile	Lys	TAA * 895		26	685

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 895 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Gly Ser Pro Gly Ile His Met Thr Ser Thr Arg Leu Gln Lys Leu Leu 1 5 10 15

Glu Phe Glu Leu Pro Gly Thr Met Glu Phe Val Asn Lys Gln Phe Asn 20 25 30 -

Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro 35 40 45

Lys Tyr Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys 50 55 60

Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly 65 70 75 80

Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr 85 90 95

Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys

Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg 120 Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val 155 Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile 165 Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly 185 His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr 200 Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro 230 Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr 250 Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala 260 Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr 280 Phe Gly Gly His Asp Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met 330 Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu 360 Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu 375 Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn 395 Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg 405 Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn 425 Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe 440 Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser

Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val 470 Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 490 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 505 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 520 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 555 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 565 Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 585 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 600 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 630 635 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 680 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 695 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 730 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 760 Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val

810

Ser T	yr	Leu	Met	Asn	Ser	Met	Ile	Pro	Tyr	Glv	Val	Lvs	Ara	T 0	~1
			820				-	825	-			70	830	Leu	GIU

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys * 885 890 895

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2622 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2622

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGA Gly	TCC Ser	ATG Met	GAG Glu	TTC Phe 5	GTG Val	AAC Asn	AAG Lys	CAG Gln	TTC Phe 10	AAC Asn	TAT Tyr	AAG Lys	GAC A sp	CCT Pro 15	GTA Val	48
ASII	GIY	vai	20	11e	Ala	Tyr	He	Lys 25	Ile	Pro	Lys	Tyr	Gly 30	Gln		96
CAG Gln	Pro	GTG Val 35	AAG Lys	GCT Ala	TTC Phe	AAG Lys	ATT Ile 40	CAT His	AAC Asn	AAA Lys	ATC Ile	TGG Trp 45	GTT Val	ATT Ile	CCG Pro	144
GAA Glu	CGC Arg 50	GAT Asp	ACA Thr	TTT Phe	ACG Thr	AAC Asn 55	CCG Pro	GAA Glu	GAA Glu	GGA Gly	GAC Asp 60	TTG Leu	AAC Asn	CCG Pro	CCG Pro	192
CCG Pro 65	GAA Glu	GCA Ala	AAG Lys	CAG Gln	GTG Val 70	CCA Pro	GTT Val	TCA Ser	TAC Tyr	TAC Tyr 75	GAT Asp	TCA Ser	ACC Thr	TAT Tyr	CTG Leu 80	240
AGC Ser	ACA Thr	GAC Asp	AAC Asn	GAG Glu 85	AAG Lys	GAT Asp	AAC Asn	TAC Tyr	CTG Leu 90	AAG Lys	GGA Gly	GTG Val	ACC Thr	AAA Lys 95	TTA Leu	288
TTC Phe	GAG Glu	CGT Ar g	ATT Ile 100	TAT Tyr	TCC Ser	ACT Thr	GAC Asp	CTG Leu 105	GGC Gly	CGT Arg	ATG Met	CTG Leu	CTG Leu 110	ACC Thr	TCA Ser	336
ATC Ile	GTC Val	CGC Arg 115	GGA Gly	ATC Ile	CCA Pro	TTT Phe	TGG Trp 120	GGT Gly	GGC Gly	AGT Ser	ACC Thr	ATT Ile 125	GAC Asp	ACG Thr	GAG Glu	384
TTG Leu	AAG Lys 130	GTT Val	ATT Ile	GAC Asp	ACT Thr	AAC Asn 135	TGC Cys	ATT Ile	AAC Asn	GTG Val	ATC Ile 140	CAA Gln	CCA Pro	GAC Asp	GGT Gly	432

AG Se 14	r ry	C AG r Ar	A TC g Se	T GA r Gl	A GA u Gl	A CT: u Lei 0	AA 1 12A L	C CT	C GT u Va	A AT 1 I1 15	e Il	CC GC le Gl	G CO	CC T	CC er	GCG Ala 160	,	480
GA: As _i	C AT	r AT e Il	C CA e Gl	G TT n Pho 16	e GII	G TGC u Cys	AAC Lys	S AG	TT r Ph	e Gl	C CA y Hi	C GA s Gl	A GT u Va	l L	TG eu 75	AAC Asn		528
CT(Let	G ACC	G CG	T AAG g Asi 180	n GI	TAC Y Tyr	GGC Gly	TCT Ser	Thi	: Gl	G TA	C AT	T CG e Ar	T TI g Ph	e Se	GC ≥r	CCA Pro		576
GA(Asp	TTC Phe	ACC Thi	Pne	GGT Gly	r TTC / Phe	GAG Glu	GAG Glu 200	Ser	CTC Let	G GAG	G GT 1 Va	T GA l As 20	p Th	C AA	AC sn	CCG Pro		624
CTC	Leu 210	, GT	C GCA / Ala	GGC Gly	AAG Lys	Phe 215	GCA Ala	ACT Thr	GA1	CCA Pro	A GC Ala 22	a Va	G AC l Th	C CI	rG eu	GCA Ala		672 ·
CAC His 225	GIU	CTG Leu	ATC Ile	CAC His	GCC Ala 230	GGT Gly	CAT His	CGT Arg	CTG Leu	TAT Tyr 235	Gly	Z ATT	r gce ≥ Ala	G AT	e .	AAC Asn 240		720
CCG Pro	AAC Asn	CGC Arg	GTG Val	TTC Phe 245	ьys	GTT Val	AAC Asn	ACC Thr	AAC Asn 250	Ala	TAC	TAC Tyr	GA(G AT Me 25	t :	AGT Ser		768
GGT Gly	TTA Leu	GAA Glu	GTA Val 260	AGC Ser	TTC Phe	GAG Glu	GAA Glu	CTG Leu 265	CGC Arg	ACG	TTC Phe	GGT Gly	GG(Gly 270	Hi	T (GAT Asp		816
GCG Ala	AAG Lys	TTT Phe 275	ATC Ile	GAC Asp	AGC Ser	TTG Leu	CAG Gln 280	GAG Glu	AAC Asn	GAG Glu	TTC	CGT Arg 285	Leu	TAC	C 7	rac Syr		864
TAC Tyr	AAC Asn 290	AAG Lys	TTT Phe	AAA Lys	GAT Asp	ATT Ile 295	GCA Ala	AGT Ser	ACA Thr	CTG Leu	AAC Asn 300	Lys	GCT Ala	AA(3 1 3 S	CC Ser		912
ATT Ile 305	GTG Val	GGT Gly	ACC Thr	ACT Thr	GCT Ala 310	TCA Ser	TTA Leu	CAG Gln	TAT Tyr	ATG Met 315	AAA Lys	AAT Asn	GTT Val	TTT Phe	: L	AA ys 20		960
GAG Glu	AAA Lys	TAT Tyr	CTC Leu	CTA Leu 325	TCT Ser	GAA Glu	Asp	Thr	Ser	GGA Gly	Lys	Phe	Ser	GTA Val 335	A	AT sp		1008
AAA Lys	TTA Leu	AAA Lys	TTT Phe 340	GAT Asp	AAG Lys	TTA Leu	Tyr	AAA Lys 345	ATG Met	TTA Leu	ACA Thr	GAG Glu	ATT Ile 350	TAC	A	CA hr		1056
GAG Glu	GAT Asp	AAT Asn 355	TTT Phe	GTT Val	AAG Lys	TTT '	TTT 1 Phe 1 360	AAA Lys	GTA Val	CTT Leu	AAC Asn	AGA Arg 365	AAA Lys	ACA Thr	T	AT yr		1104
Leu	AAT Asn 370	TTT Phe	GAT Asp	AAA Lys	Ala	GTA 1 Val 1 375	ITT I	AAG Lys	ATA Ile	AAT Asn	ATA Ile 380	GTA Val	CCT Pro	AAG Lys	Gʻ Va	ΓA al	:	1152
AAT Asn 385	TAC Tyr	ACA Thr	ATA 1	Tyr .	GAT (Asp (390	GGA :	TTT 1 Phe 1	AAT (Asn)	TTA Leu	AGA Arg 395	AAT Asn	ACA Thr	AAT Asn	TTA Leu	A.	CA La DO	1	1200
GCA . Ala .	AAC Asn	TTT . Phe	Asn	GGT Gly (CAA /	AAT A Asn 1	ACA C	ilu :	ATT . Ile .	AAT . Asn .	AAT Asn	ATG Met	AAT Asn	TTT Phe 415	A(CT ir	1	248

AAA Lys	CTA Leu	AAA Lys	AAT Asn 420	TTT Phe	ACT Thr	GGA Gly	TTG Leu	TTT Phe 425	GAA Glu	TTT	TAT Tyr	AAG Lys	TTG Leu 430	CTA Leu	TGT Cys	. 1	1296
GTA Val	AGA Arg	GGG Gly 435	ATA Ile	ATA Ile	ACT Thr	TCT Ser	AAA Lys 440	ACT Thr	AAA Lys	TCA Ser	TTA Leu	GAT Asp 445	Lys	GGA Gly	TAC Tyr	1	1344
AAT Asn	AAG Lys 450	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 455	TGT Cys	ATC Ile	AAA Lys	GTT Val	AAT Asn 460	AAT Asn	TGG Trp	GAC Asp	TTG Leu	1	L392
TTT Phe 465	TTT Phe	AGT Ser	CCT Pro	TCA Ser	GAA Glu 470	GAT Asp	AAT Asn	TTT Phe	ACT Thr	AAT Asn 475	GAT Asp	CTA Leu	AAT Asn	AAA Lys	GGA Gly 480	1	1440
GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 485	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 490	GCA Ala	GCA Ala	GAA Glu	GAA Glu	AAT Asn 495	ATT Ile	1	.488
						CAA Gln										1	.536
GAA Glu	CCT Pro	GAA Glu 515	AAT Asn	ATT Ile	TCA Ser	ATA	GAA Glu 520	AAT Asn	CTT Leu	TCA Ser	AGT Ser	GAC Asp 525	ATT Ile	ATA Ile	GGC Gly	1	.584
						AAT Asn 535										. 1	.632
						ACT Thr											.680
TTT Phe	GAA Glu	CAT His	GGT Gly	AAA Lys 565	TCT Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 570	ACA Thr	AAT Asn	TCT Ser	GTT Val	AAC Asn 575	GAA Glu	1	.728
						CGT Arg										1	. 77 6
		Lys	Val	Asn	Lys	GCT Ala	Thr	Glu	Ala	Ala	Met	Phe				. 1	.824
						GAT Asp 615										1	. 87 2
ACT Thr 625	ACG Thr	GAT Asp	AAA Lys	ATT Ile	GCG Ala 630	GAT Asp	ATA Ile	ACT Thr	ATA Ile	ATT Ile 635	ATT Ile	CCA Pro	TAT Tyr	ATA Ile	GGA Gly 640	1	.920
						AAT Asn										1	.968
GCT Ala	TTA Leu	ATA Ile	TTT Phe 660	TCA Ser	GGA Gly	GCT Ala	GTT Val	ATT Ile 665	CTG Leu	TTA Leu	GAA Glu	TTT Phe	ATA Ile 670	CCA Pro	GAG Glu	2	2016
						GGT Gly										2	2064

AA: Ası	T AA(n Ly: 69(, va.	CTA L Leu	ACC Thr	GT1	CAA Glr 699	r T111	A ATA	A GAT	AA1 Asr	GCT n Ala 700	a Lei	A AG	T AA	A AGA s Arg	2112
AA1 Asr 705	1 61	A AAA 1 Lys	TGG Trp	GAT Asp	GAG Glu 710	val	TAT	AA. Lys	TAT	11e 715	e Val	ACI Thi	A AA: Asi	r TG(TTA Leu 720	2160
GCA Ala	AAC Lys	GTI Val	AAT Asn	ACA Thr 725	GID	ATT	GAT Asp	CTA Leu	ATA Ile 730	Arg	AAA Lys	AAA Lys	ATC Met	AA/ Lys 735	GAA Glu	2208
Ата	Leu	GIU	740	Gin	Ala	Glu	Ala	745	Lys	Ala	Ile	Ile	750	Tyr	CAG Gln	2256
TYL	ASII	755	lyt	inr	Gia	GIU	760	Lys	Asn	Asn	Ile	Asn 765	Phe	Asn		2304
Asp	770	Deu	AGT Ser	ser	цуs	775	Asn	GIu	Ser	Ile	Asn 780	Lys	Ala	Met	Ile	2352
785	116	ASI	AAA Lys	Pne	790	Asn	Gin	Cys	Ser	Val 795	Ser	Tyr	Leu	Met	Asn 008	2400
ser	Mec	ile	CCT Pro	805	GIŸ	vai	rys	Arg	Leu 810	Glu	Asp	Phe	Asp	Ala 815	Ser	2448
Leu	Lys	Asp	GCA Ala 820	Leu	Leu	Lys	Tyr	11e 825	Tyr	Asp	Asn	Arg	Gly 830	Thr	Leu	2496
ATT Ile	GGT Gly	CAA Gln 835	GTA Val	GAT Asp	AGA Arg	Leu	AAA Lys 840	GAT Asp	AAA Lys	GTT Val	AAT Asn	AAT Asn 845	ACA Thr	CTT Leu	AGT Ser	2544
inr	GAT Asp 850	ATA Ile	CCT '	TTT Phe	GIn .	CTT Leu 855	TCC . Ser	AAA Lys	TAC (Val .	GAT . Asp . 860	AAT Asn	CAA Gln	AGA Arg	TTA Leu	2592
TTA Leu 865	TCT Ser	ACA Thr	TTT I	Chr (GAA ' Glu ' 870	TAT :	ATT . Ile :	AAG Lys	TAA *							2622

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 874 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Gly Ser Met Glu Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val

Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Lys Tyr Gly Gln Met 30

Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro

Glu	Arg 50	Asp	Thr	Phe	Thr	Asn 55	Pro	Glu	Glu	Gly	Asp 60	Leu	Asn	Pro	Pro
Pro 65	Glu	Ala	Lys	Gln	Val 70	Pro	Val	Ser	Tyr	Tyr 75	Asp	Ser	Thr	Tyr	Leu 80
Ser	Thr	Asp	Asn	Glu 85	Lys	Asp	Asn	Tyr	Leu 90	Lys	Gly	Val	Thr	Lys 95	Leu
Phe	Glu	Arg	Ile 100	Tyr	Ser	Thr	Asp	Leu 105	Gly	Arg	Met	Leu	Leu 110	Thr	Ser
Ile	Val	Arg 115	Gly	Ile	Pro	Phe	Trp 120	Gly	Gly	Ser	Thr	Ile 125	Asp	Thr	Glu
Leu	Lys 130	Val	Ile	Asp	Thr	Asn 135	Cys	Ile	Asn	Val	Ile 140	Gln	Pro	Asp	Gly
Ser 145	Tyr	Arg	Ser	Glu	Glu 150	Leu	Asn	Leu	Val	Ile 155	Ile	Gly	Pro	Ser	Ala 160
Asp	Ile	Ile	Gln	Phe 165	Glu	Cys	Lys	Ser	Phe 170	Gly	His	Glu	Val	Leu 175	Asn
Leu	Thr	Arg	Asn 180	Gly	Tyr	Gly,	Ser	Thr 185	Gln	Tyr	Ile	Arg	Phe 190		Pro
Asp	Phe	Thr 195	Phe	Gly	Phe	Glu	Glu 200	Ser	Leu	Glu	Val	Asp 205	Thr	Asn	Pro
Leu	Leu 210	Gly	Ala	Gly	Lys	Phe 215	Ala	Thr	Asp	Pro	Ala 220	Val	Thr	Leu	Ala
His 225	Glu	Leu	Ile	His	Ala 230	Gly.	His	Arg	Leu	Tyr 235	Gly	Ile	Ala	Ile	Asn 240
Pro	Asn	Arg	Val	Phe 245	Lys	Val	Asn	Thr	Asn 250	Ala	Tyr	Tyr	Glu	Met 255	Ser
Gly	Leu	Glu	Val 260	Ser	Phe	Glu	Glu	Leu 265	Arg	Thr	Phe	Gly	Gly 270	His	Asp
Ala	Lys	Phe 275	Ile	Asp	Ser	Leu	Gln 280	Glu	Asn	Glu	Phe	Arg 285	Leu	Tyr	Tyr
Tyr	Asn 290	Lys	Phe	Lys	Asp	11e 295		Ser	Thr	Leu	Asn 300		Ala	Lys	Ser
11e 305	Val	Gly	Thr	Thr	Ala 310	Ser	Leu	Gln	Tyr	Met 315	Ŀуs	Asn	Val	Phe	Lys 320
Glu	Lys	Tyr	Leu	Leu 325	Ser	Glu	Asp	Thr	Ser 330	Gly	Lys	-Phe	Ser	Val 335	Asp
Lys	Leu	Lys	Phe 340	Asp	Lys	Leu	Tyr	Lys 345	Met	Leu	Thr	Glu	Ile 350	Tyr	Thr
Glu	Asp	Asn 355	Phe	Val	Lys	Phe	Phe 360	Lys	Val	Leu	Asn	Arg 365	Lys	Thr	Tyr
Leu	Asn 370	Phe	Asp	Lys	Ala	Val 375	Phe	Lys	Ile	Asn	Ile 380	Val	Pro	Lys	Val
Asn 385	Tyr	Thr	Ile	Tyr	Asp 390	Gly	Phe	Asn	Leu	Arg 395	Asn	Thr	Asn	Leu	Ala 400

Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys 425 Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly 475 Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile 490 Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys 535 Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu 570 Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser 620 Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly 630 Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu 665 Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala 680 Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln

745

Tyr	Asn	Gln 755	Tyr	Thr	Glu	Glu	Glu 760	Lys	Asn	Asn	Ile	Asn 765	Phe	Asn	Ile	
Asp	Asp 770	Leu	Ser	Ser	Lys	Leu 775	Asn	Glu	Ser	Ile	Asn 780	Lys	Ala	Met	Ilę	
Asn 785	Ile	Asn	Lys	Phe	Leu 790	Asn	Gln	Cys	Ser	Val 795		Tyr	Leu	Met	Asn 800	
Ser	Met	Ile	Pro	Tyr 805	Gly	Val	Lys	Arg	Leu 810	Glu	Asp	Phe	Asp	Ala 815	Ser	
Leu	Lys	Asp.	Ala 820	Leu	Leu	Lys	Tyr	Ile 825		Asp	Asn	Arg	Gly 830	Thr	Leu	
Ile	Gly	Gln 835	Val	Asp	Arg	Leu	Lys 840	Asp	Lys	Val	Asn	Asn 845	Thr	Leu	Ser	
Thr	Asp 850	Ile	Pro	Phe	Gln	Leu 855	Ser	Lys	Tyr-	Val	Asp 860	Asn	Gln	Arg	Leu	
Leu 865	Ser	Thr	Phe	Thr	Glu 870	Tyr	Ile	Lys	*	٠					-	
(2)	INFO	RMAT	NOI	FOR	SEQ	ID N	10: 7	7: .								
	(i)	(<i>I</i> (I	A) LE 3) TY C) ST	ENGTI (PE : [RAN]	i: 26 nucl DEDNE	TERI 13 h Leic ESS: line	ase acid doub	pai:	rs					-		
	(ii)	MOI	LECUI	LE TY	(PE:	DNA	(ger	nomic	2)							
	(ix)		A) NA	ME/I	ŒY: ION: I	CDS	513									
						PTIC										
ATG Met 1	CCA Pro	TTT Phe	GTT Val	AAT Asn 5	AAA Lys	CAA Gln	TTT Phe	AAT Asn	TAT Tyr 10	AAA Lys	GAT Asp	CCT Pro	GTA Val	AAT Asn 15	GGT Gly	48
GTT Val	GAT Asp	ATT Ile	GCT Ala 20	TAT Tyr	ATA Ile	AAA Lys	ATT Ile	CCA Pro 25	AAT Asn	GCA Ala	GGA Gly	CAA Gln	ATG Met 30	CAA Gln	CCA Pro	96
GTA Val	AAA Lys	GCT Ala 35	TTT Phe	AAA Lys	ATT Ile	CAT His	AAT Asn 40	AAA Lys	AȚA Ile	TGG Trp	GTT Val	ATT Ile 45	CCA Pro	GAA Glu	AGA Arg	144
GAT Asp	ACA Thr 50	TTT Phe	ACA Thr	AAT Asn	CCT Pro	GAA Glu 55	GAA Glu	GGA Gly	GAT Asp	TTA Leu	AAT Asn 60	CCA Pro	CCA Pro	CCA Pro	GAA Glu	192
GCA Ala 65	AAA Lys	CAA Gln	GTT Val	CCA	GTT Val 70	TCA Ser	TAT Tyr	TAT Tyr	GAT Asp	TCA Ser 75	ACA Thr	TAT Tyr	TTA Leu	AGT Ser	ACA Thr 80	240
GAT	AAT	GAA Glu	AAA	GAT	AAT	TAT	TTA	AAG	GGA	GTT	ACA	AAA	TTA	TTT	GAG	288

AG/ Arg	A ATT	TATE	TCA Sei 100	Thi	GAT Asp	CTI Leu	GGA Gly	AGA Arg 105	g Met	TTC Leu	TTA Lev	A ACI	A TC	r Il	A GTA e Val	336
AGC Arg	G GG# G Gly	A ATA	Pro	TTT Phe	TGG	GGT Gly	GGA Gly 120	Ser	ACA Thr	ATA Ile	GAT RSF	ACA Thi 125	Gl	A TT	A AAA u Lys	384
GTT Val	Ile 130	: Asp	ACT Thr	'AAT 'Asn	TGT Cys	Ile 135	Asn	GTG Val	ATA Ile	CAA Gln	CCA Pro 140	Asp	GGT Gly	r AG: ⁄ Se:	TAT Tyr	432
AGA Arg 145	Ser	GAA Glu	GAA Glu	CTT Leu	AAT Asn 150	Leu	GTA Val	ATA	ATA Ile	GGA Gly 155	Pro	TCA Ser	GCT Ala	GA1	T ATT D Ile 160	480
ATA Ile	CAG Gln	TTT Phe	GAA Glu	TGT Cys 165	Lys	AGC Ser	TTT	GGA Gly	CAT His 170	Glu	GTT Val	TTG Leu	AAT	CT1 Leu 175	ACG	528
CGA Arg	AAT Asn	GGT Gly	TAT Tyr 180	GGC	TCT Ser	ACT Thr	CAA Gln	TAC Tyr 185	ATT	AGA Arg	TTT Phe	AGC Ser	CCA Pro 190	Asp	TTT Phe	576
ACA Thr	TTT Phe	GGT Gly 195	TTT Phe	GAG Glu	GAG Glu	TCA Ser	CTT Leu 200	GAA Glu	GTT Val	GAT Asp	ACA Thr	AAT Asn 205	CCT Pro	CTT Leu	TTA Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAA Lys	TTT Phe	GCT Ala	ACA Thr 215	GAT Asp	CCA Pro	GCA Ala	GTA Val	ACA Thr 220	TTA Leu	GCA Ala	CAT His	GAA Glu	672
CTT Leu 225	ATA Ile	CAT His	GCT Ala	GGA Gly	CAT His 230	AGA Arg	TTA Leu	TAT Tyr	GGA Gly	ATA Ile 235	GCA Ala	ATT Ile	AAT Asn	CCA Pro	AAT Asn 240	720
AGG Arg	GTT Val	TTT Phe	AAA Lys	GTA Val 245	TAA neA	ACT Thr	AAT Asn	GCC Ala	TAT Tyr 250	TAT Tyr	GAA Glu	ATG Met	AGT Ser	GGG Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTT Phe 260	GAG Glu	GAA Glu	CTT Leu	AGA Arg	ACA Thr 265	TTT Phe	GGG Gly	GGA Gly	CAT His	GAT Asp 270	GCA Ala	AAG Lys	816
TTT Phe	ATA Ile	GAT Asp 275	AGT Ser	TTA Leu	CAG Gln	GAA Glu	AAC Asn 280	GAA Glu	TTT Phe	CGT Arg	CTA Leu	TAT Tyr 285	TAT Tyr	TAT Tyr	AAT Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATA Ile	GCA Ala	AGT Ser 295	ACA Thr	CTT Leu	AAT Asn	AAA Lys	GCT Ala 300	AAA Lys	TCA Ser	ATA Ile	GTA Val	912
GGT Gly 305	ACT Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA . Lys :	Met :	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	Tyr	ACA Thr 350	GAG Glu	GAT A sp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	Lys	GTA (Val : 360	CTT Leu	AAC . Asn .	AGA . Arg	Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104

TTT Phe	GAT Asp 370	Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	TTE	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	Lys	GŢA Val	AA1 Asr	TAC	1152
385	116	lyr	Asp	GIY	390	Asn	Leu	Arg	Asn	395	Asn	Leu	Ala	Ala	AAC Asn 400	1200
Pne	Asn	GIY	GIN	405	THE	GIU	lle	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415		1248
гÀг	Asn	rne	420	GIY	rea	Pne	GIU	425	Tyr	Lys	Leu	Leu	Cys 430	Val	AGA Arg	1296
GIÀ	lle	11e 435	Thr	Ser	гЛs	ACT Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	Lys	1344
GCA Ala	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	AAT Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392
AGT Ser 465	CCT Pro	TCA Ser	GAA Glu	GAT Asp	AAT Asn 470	TTT Phe	ACT Thr	AAT Asn	GAT Asp	CTA Leu 475	AAT Asn	AAA Lys	GGA Gly	GAA Glu	GAA Glu 480	1440
Ile	Thr	Ser	Asp	Thr 485	Asn	ATA Ile	Glu	Ala	Ala 490	Glu	Glu	Asn	Ile	Ser 495	Leu	1488
Asp	Leu	Ile	Gln 500	Gln	Tyr	TAT Tyr	Leu	Thr 505	Phe	Asn	Phe	Asp	Asn 510	Glu	Pro	1536
GAA Glu	AAT Asn	ATT Ile 515	TCA Ser	ATA Ile	GAA Glu	AAT Asn	CTT Leu 520	TCA Ser	AGT Ser	GAC Asp	ATT Ile	ATA Ile 525	GGC Gly	CAA Gln	TTA Leu	1584
GAA Glu	CTT Leu 530	ATG Met	CCT Pro	AAT Asn	ATA Ile	GAA Glu 535	AGA Arg	TTT Phe	CCT Pro	AAT Asn	GGA Gly 540	AAA Lys	AAG Lys	TAT Tyr	GAG Glu	1632
TTA Leu 545	GAT Asp	AAA Lys	TAT Tyr	ACT Thr	ATG Met 550	TTC Phe	CAT	TAT Tyr	CTT Leu	CGT Arg 555	GCT Ala	CAA Gln	GAA Glu	TTT Phe	GAA Glu 560	1680
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	ATT Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 525	TTA Leu	1728
TTA Leu	AAT Asn	CCT Pro	AGT Ser 580	CGT Arg	GTT Val	TAT Tyr	Thr	TTT Phe 585	TTT Phe	TCT Ser	TCA Ser	GAC Asp	TAT Tyr 590	GTA Val	AAG Lys	1776
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT Ala	ACG Thr	GAG Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	TGG Trp	GTA Val	GAA Glu	1824
Gln	TTA Leu 610	GTA Val	TAT Tyr	GAT Asp	TTT Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr	1872
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	GAT Asp	ATA Ile 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT Tyr	ATA Ile	GGA Gly	CCT Pro	GCT Ala 640	1920

TT Le	A AA u As	T AT	CA GO	ST AF .y As 64	in me	G TT/ t Lei	А ТАЗ	r AA.	A GAT S Asp 650	O As	T TT	T GT.	A GG l Gl	T GC y Al 65	T TTA a Leu 5		1968
AT.	A TT e Ph	T TC e Se	A GG r Gl 66	À YI	T GT a Va	T ATT	CTC Leu	Let 669	ı Glu	A TT:	T ATA	A CCA	A GA G Gl: 67	u Il	T GCA e Ala		2016
ATA Ile	A CC	F GT Va 67	1 Le	A GG u Gl	T AC	r TTI r Phe	GCA Ala 680	Leu	r GTA ı Val	TC# Ser	TAT	AT1 : Ile : 685	e Ala	G AA a As	T AAG n Lys		2064
GT7 Val	CTA Leu 690	ı Th	C GT r Va	T CA l Gl	A ACI	ATA Ile 695	Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	Lys	AGA Arc	A AA' J Asi	r GAA n Glu		2112
AAA Lys 705	Trp	GA' As	r GA	G GT	TAT Tyr 710	Lys	TAT Tyr	ATA Ile	GTA Val	ACA Thr 715	Asn	TGG Trp	Leu	GC/ Ala	A AAG A Lys 720		2160
GTT Val	AAT Asn	ACA Th:	A CAG	729	a Asp	CTA Leu	ATA Ile	AGA Arg	AAA Lys 730	Lys	ATG Met	AAA Lys	GAA Glu	GCT Ala 735	TTA Leu		2208
GAA Glu	AAT Asn	CA/ Glr	A GCA 1 Ala 740	Glu	A GCA 1 Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAI -Tyr	AAT Asn	٠	2256
GIn	Tyr	755	Glu	Glu	Glu	AAA Lys	760	Asn	Ile	Asn	Phe	Asn 765	Ile	Asp	Asp		2304
Leu	770	Ser	Lys	Leu	Asn	GAG Glu 775	Ser	Ile	Asn	Lys	Ala 780	Met	Ile	Asn	Ile		2352
785	Lys	Phe	Leu	Asn	Gln 790		Ser	Val	Ser	Tyr 795	Leu	Met	Asn	Ser	Met 800		2400
11e	Pro	Tyr	Gly	Val 805	Lys	CGG Arg	Leu	Glu	Asp 810	Phe	Asp	Ala	Ser	Leu 815	Lys		2448
GAT Asp	GCA Ala	TTA Leu	Tra Leu 820	AAG Lys	TAT Tyr	ATA Ile	Tyr .	GAT Asp 825	AAT . Asn .	AGA Arg	GGA Gly	Thr	TTA Leu 830	ATT Ile	GGT Gly	•	2496
Gin	Val	835	Arg	Leu	Lys		Lys ` 840	Val .	Asn <i>I</i>	Asn '	Thr	Leu : 845	Ser	Thr	Asp		2544
ATA Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	Ser	AAA 1 Lys 1 855	TAC (Tyr 1	GTA (Val ,	GAT 1 Asp 1	Asn (CAA / Gln / 860	AGA : Arg :	ITA ' Leu '	TTA Leu	TCT Ser	:	2592
ACA Thr 865	TTT Phe	ACT Thr	GAA Glu	TAT Tyr	ATT Ile 870	AAG Lys										:	2613

(2) INFORMATION FOR SEQ ID NO: 8:

⁽i) SEQUENCE CHARACTERISTICS:

⁽A) LENGTH: 871 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Pro Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile
145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
435
440
445

Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450 460

Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu 465 470 475 480

Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490 495

Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 500 505 510

Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu 515 520 525

Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 530 535 540

Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 545 550 555 560

His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 565 570 575

Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys
580 585 590

Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu
595 600 605

Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 610 615 620

Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 625 630 635 640

Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu 645 650 655

Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 660 665 670

Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 675 680 685

Val	Leu 690	Thr	Val	Gln	Thr	Ile 695	Asp	Asn	Ala	Leu	Ser 700	Lys	Arg	Asn	Glu		
Lys 705	Trp	Asp	Glu	Val	Tyr 710	Lys	Tyr	Ile	Val	Thr 715	Asn	Trp	Leu	Ala	Lys 720		
Val	Asn	Thr	Gln	Ile 725	Asp	Leu	Ile	Arg	Lys 730	Lys	Met	Lys	Glu	Ala 735	Leu		
Glu	Asn	Gln	Ala 740	Glu	Ala	Thr	Lys	Ala 745	Ile	Ile	Asn	Tyr	Gln 750	Tyr	Asn		
Gln	Tyr	Thr 755	Glu	Glu	Glu	Lys	Asn 760	Asn	Ile	Asn	Phe	Asn 765	Ile	Asp	Asp		
Leu	Ser 770	Ser	Lys	Leu	Asn	Glu 775	Ser	Ile	Asn	Lys	Ala 780	Met	Ile	Asn	Ile		
Asn 785	Lys	Phe	Leu	Asn	Gln 790	Cys	Ser	Val	Ser	Tyr 795	Leu	Met	Asn	Ser	Met 800		
Ile	Pro	Tyr	Gly	Val 805	Lys	Arg	Leu	Glu	Asp 810	Phe	Asp	Ala	Ser	Leu 815	Lys		
Asp	Ala	Leu	Leu 820	Lys	Tyr	Ile	Tyr	Asp 825	Asn	Arg	Gly	Thr	Leu 830	Ile	Gly		
Gln	Val	Asp 835	Arg	Leu	Lys	Asp	Lys 840	Val	Asn	Asn	Thr	Leu 845	Ser	Thr	Asp		
Ile	Pro 850	Phe	Gln	Leu	Ser	Lys 855	Tyr	Val	Asp	Asn	Gln 860	Arg	Leu	Leu	Ser		
Thr 865	Phe	Thr	Glu	Tyr	Ile 870	Lys											
(2)	INF	ORMA	TION	FOR	SEQ	ID I	NO:	9:									
•	(i	() () ()		ENGT	nuc DEDN	628 leic ESS:	base aci dou	pai: d	rs								
			LECU:		YPE:	DNA	(ge	nomi	c)								
	(ix	(.	ATURI A) N. B) L	AME/	KEY : ION :	CDS	628										
	(xi) SE	QUEN	CE D	ESCR	IPTI	ON:	SEQ	ID N	0: 9	:						
ATG Met	Gln	TTC Phe	GTG Val	AAC Asn 5	AA G	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn 15	GGT Gly	4	8
GT1 Val	GAC Asp	ATT	GCC Ala 20	Tyr	ATC	AAA Lys	ATT	CCA Pro 25	Asn	GCC	GGC Gly	CAG Gln	ATG Met 30	CAG Gln	CCG Pro	9	96
GT0 Val	AAG Lys	GCT Ala 35	Phe	AAG Lys	ATT Ile	CAT His	AAC Asr	ı Lys	ATC	TGG	GTT Val	ATT Ile 45	Pro	GAA Glu	. CGC . Arg	14	14
GAT Ası	T ACA	Phe	ACG Thr	AAC Asr	CCG Pro	GAA Glu	ı Glı	A GGA u Gly	GAC Asp	TTO Lev	AAC Asr 60) Pro	CCG Pro	CCG	GAA Glu	19	₹2

GCA Ala 65	a Lys	G CA s Gl:	G GT n Va	G CC.	A GTT > Val 70	. Sei	TAC Ty	TAC TY	C GA' r As	T TC	r Th	C TA	T CI	G Ac	C AC	240
GA(Asp	AAC Asi	GA(G AAG Lys	G GA S Asi 85) Asr	TAC	CTC Lev	AA(Ly:	G GG/ S Gl ₃	/ Val	G AC l Th	C AA r Ly	A.TI s Le	u Ph	C GAC se Glu	288
CGT	ATT	TA:	TCC Ser 100	Thi	GAC Asp	CTG Leu	GGC Gly	CG7 Arg 109	g Met	CTC Lev	G CT	G AC u Th	C TC r Se ll	r Il	C GTO e Val	336
CGC	GGA Gly	ATC 116	Pro	TTT Phe	TGG	GGT Gly	GGC Gly 120	Ser	Thr	ATT	GA(C AC Th	r Gl	G TT u Le	G AAG u Lys	384
GTT Val	Ile 130	Asp	ACI Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	As	C GG p Gl	r AG Y Se	C TAC	432
AGA Arg 145	Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC	ATC Ile	GGG Gly 155	Pro	TC(Sea	GCC Ala	G GAG	C ATT D Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	Let	AA(Asr	CTC Let 175	ACG Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC	AGC Ser	CCA Pro	Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	Pro	CTG Leu	TTG Leu	624 .
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	Glu .	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT	GCA Ala	AGT . Ser ' 295	ACA Thr	CTG Leu	AAC Asn	Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	Ser	TTA (Leu (310	CAG ' Gln '	TAT . Tyr !	ATG Met	Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT (CTC Leu	CTA Leu	Ser	GAA Glu 325	GAT A	ACA S	TCT (Ser (Gly	AAA Lys 330	TTT :	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008

AA. Ly:	A TTT	GA1	AAC Lys 340	ner.	A TAC	Lys	ATG Met	Leu 349	Thr	GAC	G ATT	TAC Typ	C AC	r Gl	G GAT	105 <u>é</u>
AA? Asr	TTI Phe	GTI Val	. Lys	TT1	TTT Phe	Lys	GTA Val 360	CT1 Leu	AAC Asn	AGA Arg	AAA Lys	A ACA This	Ty	TTC Lev	G AAT 1 Asn	1104
TT7 Phe	GAT Asp 370	PAS	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	Lys	GTA Val	AA7 Asr	TAC Tyr	1152
ACA Thr 385	TTE	TAT	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC	AAT Asn	AAG Lys	1344
AGC Ser	GCT Ala 450	GAT Asp	GGG Gly	GCA Ala	TTA Leu	AAT Asn 455	GAT Asp	TTA Leu	TGT Cys	ATC Ile	AAA Lys 460	GTT Val	AAT Asn	AAT Asn	TGG Trp	1392
GAC Asp 465	ŤTG Leu	TTT Phe	TTT	AGT Ser	CCT Pro 470	TCA Ser	GAA Glu	GAT Asp	AAT Asn	TTT Phe 475	ACT Thr	AAT Asn	GAT Asp	CTA Leu	AAT Asn 480	1440
AAA Lys	GGA Gly	GAA Glu	GAA Glu	ATT Ile 485	ACA Thr	TCT Ser	GAT Asp	ACT Thr	AAT Asn 490	ATA Ile	GAA Glu	GCA Ala	GCA Ala	GAA Glu 495	GAA Glu	1488
AAT Asn	ATT Ile	AGT Ser	TTA Leu 500	GAT A sp	TTA Leu	ATA Ile	CAA Gln	CAA Gln 505	TAT Tyr	TAT Tyr	TTA Leu	ACC Thr	TTT Phe 510	AAT Asn	TTT Phe	1536
GAT Asp	AAT Asn	GAA Glu 515	CCT Pro	GAA Glu	AAT Asn	lle	TCA Ser 520	ATA Ile	GAA Glu	AAT Asn	CTT Leu	TCA Ser 525	AGT Ser	GAC Asp	ATT Ile	1584
ATA Ile	GGC Gly 530	CAA Gln	TTA Leu	GAA Glu	CTT Leu	ATG Met 535	CCT Pro	AAT Asn	ATA Ile	GAA Glu	AGA Arg 540	TTT Phe	CCT Pro	AAT Asn	GGA Gly	1632
AAA Lys 545	AAG Lys	TAT Tyr	GAG Glu	TTA Leu	GAT Asp 550	AAA Lys	TAT Tyr	ACT Thr	Met	TTC Phe 555	CAT His	TAT Tyr	CTT Leu	CGT Arg	GCT Ala 560	1680
CAA Gln	GAA Glu	TTT Phe	GAA Glu	CAT His 565	GGT Gly	AAA Lys	TCT . Ser .	Arg	ATT Ile 570	GCT Ala	TTA Leu	ACA Thr	AAT Asn	TCT Ser 575	GTT Val	1728
AAC Asn	GAA Glu	GCA Ala	TTA Leu 580	TTA Leu	AAT Asn	CCT . Pro	Ser :	CGT Arg 585	GTT Val	TAT Tyr	ACA Thr	TTT Phe	TTT Phe 590	TCT Ser	TCA Ser	1776
GAC Asp	TAT Tyr	GTA Val 595	AAG Lys	AAA Lys	GTT . Val .	Asn	AAA Lys . 000	GCT Ala	ACG Thr	GAG Glu	Ala	GCT Ala 605	ATG Met	TTT Phe	TTA Leu	1824

610 619	620	1872
625 630	GCG GAT ATA ACT ATA ATT ATT CCA TAT Ala Asp Ile Thr Ile Ile Ile Pro Tyr 635	1920
645	GGT AAT ATG TTA TAT AAA GAT GAT TTT Gly Asn Met Leu Tyr Lys Asp Asp Phe 650 655	1968
660	GGA GCT GTT ATT CTG TTA GAA TTT ATA Gly Ala Val Ile Leu Leu Glu Phe Ile 665	2016
675	TTA GGT ACT TTT GCA CTT GTA TCA TAT Leu Gly Thr Phe Ala Leu Val Ser Tyr 680 685	2064
690 695	GTT CAA ACA ATA GAT AAT GCT TTA AGT Val Gln Thr Ile Asp Asn Ala Leu Ser 700	2112
705 710 710	GAG GTC TAT AAA TAT ATA GTA ACA AAT Glu Val Tyr Lys Tyr Ile Val Thr Asn 715 720	2160
TGG TTA GCA AAG GTT AAT ACA Trp Leu Ala Lys Val Asn Thr 725	CAG ATT GAT CTA ATA AGA AAA AAA ATG Gln Ile Asp Leu Ile Arg Lys Lys Met 730 735	2208
AAA GAA GCT TTA GAA AAT CAA Lys Glu Ala Leu Glu Asn Gln 740	GCA GAA GCA ACA AAG GCT ATA ATA AAC Ala Glu Ala Thr Lys Ala Ile Ile Asn 745 750	2256
THE GIR THE ASH GIR THE	GAG GAA GAG AAA AAT AAT ATT AAT TTT Glu Glu Glu Lys Asn Asn Ile Asn Phe 760 765	2304
AAT ATT GAT GAT TTA AGT TCG ASN Ile Asp Asp Leu Ser Ser 770	AAA CTT AAT GAG TCT ATA AAT AAA GCT Lys Leu Asn Glu Ser Ile Asn Lys Ala 780	2352
ATG ATT AAT ATA AAT AAA TTT 1 Met Ile Asn Ile Asn Lys Phe I 785	TTG AAT CAA TGC TCT GTT TCA TAT TTA Leu Asn Gln Cys Ser Val Ser Tyr Leu 795 800	2400
ATG AAT TCT ATG ATC CCT TAT (Met Asn Ser Met Ile Pro Tyr (805	GT GTT AAA CGG TTA GAA GAT TTT GAT Hy Val Lys Arg Leu Glu Asp Phe Asp 810 815	2448
GCT AGT CTT AAA GAT GCA TTA 1 Ala Ser Leu Lys Asp Ala Leu 1 820	TA AAG TAT ATA TAT GAT AAT AGA GGA eu Lys Tyr Ile Tyr Asp Asn Arg Gly 825	2496
I'm bed lie Gly Gin val Asp A	GA TTA AAA GAT AAA GTT AAT AAT ACA rg Leu Lys Asp Lys Val Asn Asn Thr 40 845	2544
CTT AGT ACA GAT ATA CCT TTT C Leu Ser Thr Asp Ile Pro Phe G 850 855	AG CTT TCC AAA TAC GTA GAT AAT CAA ln Leu Ser Lys Tyr Val Asp Asn Gln 860	2592
AGA TTA TTA TCT ACA TTT ACT G Arg Leu Leu Ser Thr Phe Thr G 865	AA TAT ATT AAG TAA lu Tyr Ile Lys * 875	2628

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 876 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro
20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 110

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320

Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350

Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365

Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 375 380

Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 390 395 400

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415

Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg
420 425 430

Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 435 440 445

Ser Ala Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp 450 455 460

Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn 465 470 475 480

Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu 485 490 495

Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe 500 505 510

Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile 515 520 525

Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly
530 535 540

Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala 545 550 555 560

Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val 565 570 575

Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe-Phe Ser Ser 580 585 590

Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu 595 600 605

Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu 610 615 620

Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr 625 630 635 640

Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe 645 650 655

Val	Gly	Ala	Leu 660	Ile	Phe	Ser	Gly	Ala 665	Val	Ile	Leu	Leu	Glu 670	Phe	Ile
Pro	Glu	Ile 675	Ala	Ile	Pro	Val	Leu 680	Gly	Thr	Phe	Ala	Leu 685	Val	Ser	Tyr
Ile	Ala 690	Asn	Lys	Val	Leu	Thr 695	Val	Gln	Thr	Ile	Asp 700	Asn	Ala	Leu	Ser
Lys 705	Arg	Asn	Glu	Lys	Trp 710	Asp	Glu	Val	Tyr	Lys 715	Tyr	Ile	Va1	Thr	Asn 720
Trp	Leu	Ala	Lys	Val 725	Asn	Thr	Gln	Ile	Asp 730	Leu	Ile	Arg	Lys	Lys 735	Met
Lys	Glu	Ala	Leu 740	Glu	Asn	Gln	Ala	Glu 745	Ala	Thr	Lys	Ala	Ile 750	Ile	Asn
Ţyr	Gln	Tyr 755	Asn	Gln	Tyr	Thr	Glu 760	Glu	Glu	Lys	Asn	Asn 765	Ile	Asn	Phe
Asn	Ile 770	Asp	Asp	Leu	Ser	Ser 775	Lys	Leu	Asn	Glu	Ser 780	Ile	Asn	Lys	Ala
Met 785	Ile	Asn	Ile	Asn	Lys 790	Phe	Leu	Asn	Gln	Cys 795	Ser	Val	Ser	туг	Leu 800
Met	Asn	Ser	Met	Ile	Pro	Tyr	Gly	Val	Lys	Arg	Leu	Glu	Asp	Phe	Asp

Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly 820

810

Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr

Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln

Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 870

(2) INFORMATION FOR SEQ ID NO: 11:

805

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2637 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE_TYPE: DNA (genomic)
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2637
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT 48 Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 1

GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG 96 Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

V 41	. <u>.</u> ,	3	15	.c D	, 3 1.	IT CA le Hi	5 As	10	/S I.	ie L	rp v	al 1	11e 45	Pro	G1	.u	Arg	144
GAT Asp	ACA Thi		T AC	G AA	AC CO	CG GA CO Gl 5	<u>. G</u> 1	AA GO .u Gl	SA GA .y As	AC T	eu A	AC C sn F 60	CCG	CCG Pro	CC Pr	G o	GAA Glu	192
GCA Ala 65	Бyз	G CA	G GT n Va	G CC	U Va	T TC	A TA	C TA	C GA	p Se	CA A er T 75	CC T hr T	'AT 'yr	CTG Leu	AG Se	C i	ACA Thr 80	240
Asp	ASI		и гу	8 8	p As 5	C TAC n Ty:	. re	и њу	s G1	y Va 0	al T	hr L	ys 1	Leu	Phe 95	e (lu	288
Arg	116	ιу.	10	0	I AS	C CTO	1 GI	10	g Me 5	t Le	eu Le	eu T	hr S	Ser 10	Ile	2 V	al.	336
Arg	GIY	115	5 Pro	o Pue	e II	G GG7 p Gly	120	y Se:	r Th	r Il	e As	sp Th	nr 0 25	lu	Leu	ı L	ys	384
GTT Val	ATT Ile 130	GA(C ACT	AA 7 12A :	TG0	C ATT s Ile 135	ASI	C GTO n Val	G ATO	CA Gl	A CC n Pr 14	O As	AC G sp G	GT ly	AGC Ser	T	AC Yr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTI Leu	AAC Asr 150	CTC Leu	GTA Val	ATC . Ile	ATC Ile	GG Gl; 15	y Pr	C TC	C G	CG (GAC Asp	I.	IT Le 50	480
ATC (CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	гуѕ	AGC Ser	TTT	GGC Gly	CAC His	Gli	A GT i Va	G TT l Le	G A	sn I	CTG Leu L75	A(G ir	528
CGT A	AAC Asn	GGT Gly	TAC Tyr 180	GIY	TCT Ser	ACT	CAG Gln	TAC Tyr 185	ATT	CGT Arg	TT J Ph	C AG e Se	C CC r Pr 19	CO F	SAC Asp	TT	C le	576 ·
ACG Thr I	rne	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thi	20!	n Pr	G C	TG eu	TI Le	G u	624
GGT G	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	Thi 220	Let	G GC	A C a H	AC is	GA Gl	G u	672
CTG A Leu I 225	TC !	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCC Ala	ATT Ile	AA 7 RA 9	C C	ro	AA As: 24	n.	720
CGC G Arg V	TG :	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATC Met	AG Se	r G	GT ly 55	TT: Le	1	768
GAA G Glu V	TA A	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GA' As ₁	o A	CG /	AA(Lys	3	816
TTT A	Te t	SAC Asp 275	AGC Ser	TTG Leu	CAG Gln	Glu .	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	Ty	TA Ty	AC I	AA(12	:	864
AAG T Lys Pl 2	TT A he L 90	'Àa 'YY	GAT Asp	ATT Ile	GCA Ala	AGT A Ser 1 295	ACA Thr	CTG Leu	AAC Asn	AA G Lys	GCT Ala 300	AAG Lys	TC(Ser	AT Il	e V	GTO /al	1	912

GG1 G1 305		ACT	GCT Ala	TCA Ser	TTA Leu 310	GIL	TAT Tyr	ATC	G AAA	AAAA Asr 315	ı Val	r TT: L Phe	Γ AA ∋ Ly	A GA	G AAA Lys 320	960
TAI Tyr	CTC	CTA Leu	TCI Ser	GAA Glu 325	Asp	ACA Thr	TCT	GGZ Gly	A AAA / Lys 330	Phe	TCC Ser	GTA Val	GA'	T AA p Ly: 33!	A TTA 5 Leu 5	1008
Буз	FIIC	, wah	340	Leu	IYE	пÀг	met	145 345	Thr	Glu	Ile	Tyr	350	r Glu	GAT Asp	1056
ASII	FIIC	355	цуѕ	File	Pne	rys	360	Leu	Asn	Arg	Lys	Thr 365	Ту	Let	AAT Asn	1104
	370	пуs	ALA	Vai	rne	375	ile	Asn	Ile	Val	9ro 380	Lys	Val	. Asn	TAC	1152
385		-7-	vab	GIA	390	ASII	Leu	Arg	Asn	Thr 395	Asn	Leu	Ala	Ala	AAC Asn 400	1200
2116	ASII	GIY	GIN	405	inr	GTA	ile	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415		1248
Буз	Asii	FILE	ACT Thr 420	GTA	reu	Pne	Glu	Phe 425	Tyr	Lys	Leu	Leu	Cys 430	Val	Arg	1296
GIY	116	435	ACT Thr	ser	гÀг	Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	Lys	1344
116	450	GIY	CGT Arg	cys	Asp	455	Ala	Leu	Asn	Asp	Leu 460	Cys	Ile	Lys	Val	1392
465	.NoII	rrp	GAC Asp	rea	470	Pue	Ser	Pro	Ser	Glu 475	Asp	Asn	Phe	Thr	Asn 480	1440
rap	Den	ASII		485	GIU	GIU	lie	Thr	Ser 490	Asp	Thr	Asn	Ile	Glu 495	Ala	1488
AId	GIU	GIU	AAT Asn 500	ite	ser	Leu	Asp	Leu 505	Ile	Gln	Gln	Tyr	Tyr 510	Leu	Thr	1536
FIIE	ASN	515	GAT Asp	ASN	Glu	Pro	Glu 520	Asn	Ile	Ser	Ile	Glu 525	Asn	Leu	Ser	1584
Ser	530	TIE	ATA Ile	GIY	GIn	Leu 535	Glu	Leu	Met	Pro .	Asn 540	Ile	Glu	Arg	Phe	1632
545	ASII	GIY	AAA Lys	rys	550	GIU	Leu .	Asp	Lys	Tyr ' 555	Thr	Met	Phe	His	Tyr 560	1680
CTT Leu	CGT Arg	GCT Ala	CAA (Gln (GAA Glu 565	TTT (Phe (GAA Glu	CAT (GGT Gly	AAA Lys 570	TCT . Ser .	AGG Arg	ATT Ile	GCT Ala	TTA Leu 575	ACA Thr	1728

•					٠.	•			
AAT TCT Asn Ser	GTT AF Val As 58	Gra A	CA TTA 1 la Leu I	TA AAT eu Asn 585	Pro Sei	r CGT GT r Arg Va	T TAT ACA 1 Tyr Thr 590	TTT	1776
TTT TCT Phe Ser	TCA GA Ser As 595	C TAT G	ar nas n	AA GTT ys Val 00	AAT AAA Asn Lys	A GCT ACC S Ala Thi 605	G GAG GCA Glu Ala	GCT Ala	1824
ATG TTT Met Phe 610	TTA GG Leu Gl	C TGG G y Trp Va	TA GAA C al Glu G 615	AA TTA ln Leu	GTA TAT Val Tyr	GAT TTT Asp Phe 620	ACC GAT Thr Asp	GAA Glu	1872
ACT AGC Thr Ser 625	GAA GTA Glu Va	A AGT AG 1 Ser Th 63	T THE M	AT AAA sp Lys	ATT GCG Ile Ala 635	Asp Ile	ACT ATA Thr Ile	ATT Ile 640	1920
ATT CCA Ile Pro	TAT ATA	A GGA CC e Gly Pr 645	T GCT T	u Asn	ATA GGT Ile Gly 650	AAT ATG Asn Met	TTA TAT Leu Tyr 655	AAA Lys	1968
	660	dly Ri	a neu li	665	ser Gly	Ala Val	ATT CTG Ile Leu 670	Leu	2016
	ATA CCA Ile Pro 675	GAG AT	T GCA AT e Ala Il 68	e Pro	GTA TTA Val Leu	GGT ACT Gly Thr 685	TTT GCA Phe Ala	CTT Leu	2064
GTA TCA 1 Val Ser 1 690	TAT ATT Tyr Ile	GCG AA Ala Asi	T AAG GT n Lys Va 695	T CTA I	ACC GTT Thr Val	CAA ACA Gln Thr 700	ATA GAT	AAT Asn	2112
GCT TTA A Ala Leu S 705	GT AAA Ger Lys	AGA AAT Arg Asr 710	r Gru ry	A TGG C	GAT GAG Asp Glu 715	GTC TAT Val Tyr	Lys Tyr	ATA Ile 720	2160
GTA ACA A Val Thr A	AT TGG sn Trp	TTA GCA Leu Ala 725	AAG GT Lys Val	L Asn T	CA CAG Chr Gln	ATT GAT Ile Asp	CTA ATA A Leu Ile A 735	AGA Arg	2208
AAA AAA A Lys Lys M	TG AAA et Lys 740	GAA GCT Glu Ala	TTA GAA Leu Glu	AAT C Asn G 745	AA GCA (ln Ala (Glu Ala :	ACA AAG C Ihr Lys A 750	CT la	2256
ATA ATA A Ile Ile A 7	AC TAT sn Tyr 55	CAG TAT Gln Tyr	AAT CAA Asn Gln 760	TAL I	CT GAG (hr Glu (GAA GAG A Glu Glu I 765	Tha Yeu Y	AT sn	2304
ATT AAT T Ile Asn P 770	TT AAT he Asn	ATT GAT Ile Asp	GAT TTA Asp Leu 775	AGT TO Ser Se	er Lys L	ITT AAT C eu Asn C	GAG TCT A Slu Ser I	TA le	2352
AAT AAA GO Asn Lys Al 785	T ATG	ATT AAT Ile Asn 790	ATA AAT Ile Asn	AAA TT	TT TTG A ne Leu A 795	AT CAA T sn Gln C	ys Ser V	TT al 00	2400
TCA TAT TT Ser Tyr Le	u Met 1	AAT TCT Asn Ser 805	ATG ATC Met Ile	CCT TA	r Gly V	TT AAA C al Lys A	GG TTA GA rg Leu GI 815	AA Lu	2448
GAT TTT GA Asp Phe As	T GCT A P Ala S 820	AGT CTT Ser Leu	AAA GAT Lys Asp	GCA TT Ala Le 825	TA TTA A	ys Tyr I	TA TAT GA le Tyr As 30	AT Sp	2496
AAT AGA GG Asn Arg Gl 83	7 **** 1	TTA ATT Leu Ile	GGT CAA Gly Gln 840	GTA GA Val As	T AGA T p Arg Le	TA AAA G eu Lys A: 845	AT AAA GI sp Lys Va	T	2544

AAT ACA CTT AGT ACA GAT ATA CCT TTT CAG CTT TCC AAA TAC GTA 2592 Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 855 860 GAT AAT CAA AGA TTA TTA TCT ACA TTT ACT GAA TAT ATT AAG TAA 2637 Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys * 875

- (2) INFORMATION FOR SEQ ID NO: 12:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 879 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12: Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 105 Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 230

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 475 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 490 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 505 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 520 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala

- Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 620
- Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 640
- Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
 645 650 655
- Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670
- Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685
- Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700
- Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
 705 710 715 720
- Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
 725 730 735
- Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
 740 745 750
- Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765
- Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 770 780
- Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800
- Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815
- Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830
- Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845
- Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
- Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 865 870 875
- (2) INFORMATION FOR SEQ ID NO: 13:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2862 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2862
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

1	l Gr		ic ve	,	5	AG CA /s Gl	.n Pi	ie A	sn 1	yr L	ys A	sp 1	Pro	Va:	l As	sn 15	Gly	48
GTT Val	GA0	C AI	. C A1	C TA a Ty	C AT	C AA .e Ly	A AT	e Pi	CA AA ro As 25	AC GO	CC G la G	GC (CAG Sln	ATO Met	: G1	lG .n	CCG Pro	96
GTG Val	Lys	, 'V'	T TT a Ph 5	C AA e Ly	G AT s Il	T CA e Hi	S AS	C AA in Ly 0	NA AT	rc to Le Ti	G G	TT # al I	TT le 45	CCG	GA Gl	A u	CGC Arg	144
GAT Asp	ACA Thr 50		T AC e Th	G AA	c cc n Pr	G GA o Gl: 5	n GT	A GG u Gl	A GA y As	C TI	u A	AC C sn P	CG TO	CCG Pro	CC Pr	G (GAA Glu	192
GCA Ala 65	Lys	Gli	G GT(n Va	G CCI l Pro	A GT O Va. 7	T TC l Se 0	A TA	C TA	C GA r As	p Se	A AC r Th 5	CC T	AT yr	CTG Leu	AG Se:	C 1 r 1	ACA Thr 80	240
GAC Asp	AAC Asn	GA0	G AAG	G GAT S Asp 89	ASI	TAC Tyr	CTC	G AA	G GG. s Gl [.] 9	y Va	G AC	C A	AA ys	TTA Leu	TTO Pho	9 (SAG Slu	288
CGT Arg	ATT Ile	TAT Tyr	TCC Ser 100		GAC Asp	CTC Leu	GGG Gly	C CG	g Mei	G CT	G CT u Le	G Ac	nr :	TCA Ser 110	ATO	C G	TC al	336
CGC Arg	GGA Gly	ATC Ile 115	PLO	TTT Phe	Trp	GGT Gly	GGC Gly 120	' Ser	T ACC	C AT	ΓGA e As	Tr q	CG (GAG Glu	TTC Leu	, A	AG ys	384
Val	ATT Ile 130	GAC As p	ACT Thr	AAC Asn	TGC Cys	Ile 135	AAC Asn	GTG Val	ATO	CAA Glr	A CC. 1 Pro 14	o As	p G	GT Sly	AGC Ser	T	AC yr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	Pro	C TC	C G	CG la	GAC Asp	I	TT le 50	480
ATC (CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	Glu	GTC Val	TT Le	G A u A	sn :	CTG Leu 175	A(Ti	CG nr	528
CGT A	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	He	CGT Ar g	Phe	AG Se:	r P	ro A	GAC Asp	T7 Pł	rc ie	576
ACG Thr I	- IIC	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAG Asi 205	a P	CG (ro I	CTG Leu	TI Le	G u	624
GGT C Gly A	GCA (Ala (210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	Let	G GC	CA C la E	CAC	GA G1	.G u	672
CTG A Leu I 225	ATC (CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT	AA As	AC C	CG To	AA As 24	n	720
CGC G Arg V	TG 7	TTC Phe	гÀг	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AG Se	r G	GT ly 55	TT: Le:	A u	768
GAA G Glu V	TA A	er	TTC Phe 260	GAG (Glu (GAA Glu	CTG (Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GA As 27	p A	CG .	AA(Ly:	3	816

TT	T ATO	C GAG Asj 27	p oc.	TT(G CAG	GAG Glu	AAC Asn 280	r GT	TTO Phe	C CG?	r cro	G TAC 1 Ty: 28	r Ty	C TA	C AAC		864
AA(Lys	3 TT	- <i>-</i>	A GAT s Asp	T ATT	GCA Ala	AGT Ser 295	inr	CTC Lev	AAC Asn	Lys	GCT Ala 300	ı Lys	G TC	C AT	GTG Val		912
305	5	. 1111	. Alc	Ser	310	GIN	Tyr	Met	Lys	315	ı Val	. Phe	Lys	Glu	AAA Lys 320		960
	. nec	. Dec	, ser	325	д ж	inr	ser	GIY	330	Phe	Ser	· Val	Asp	335		10	800
_,		, NSP	340	Deu	TYL	гуѕ	Mec	345	Inr	Glu	Ile	Tyr	350	Glu	GAT Asp	10	056
ASI	FILE	355	. шуз	PHE	Pne	rys	360	Leu	Asn	Arg	Lys	Thr 365	Tyr	Leu	AAT Asn	11	L04
FIIC	370	Lys	Ala	vai	Pne	175 375	11e	Asn	Ile	Val	9ro 380	Lys	Val	Asn	_	11	.52
385	. 116	TYL	мар	GIĄ	TTT Phe 390	Asn	Leu	Arg	Asn	Thr 395	Asn	Leu	Ala	Ala	Asn 400	12	200
rne	ASII	GIĄ	GIN	405	ACA Thr	GIu	lie	Asn	Asn 410	Met	Asn	Phe	Thr	Lys 415	Leu	12	48
. Lys	ASII	PHE	420	GIY	TTG Leu	Pne	Glu	Phe 425	Tyr	Lys	Leu	Leu	Cys 430	Val	Arg	12	96
GIÀ	TIE	435	inr	ser	AAA Lys	Thr	Lys 440	Ser	Leu	Asp	Lys	Gly 445	Tyr	Asn	Lys	13	44
116	450	GIY	Arg	Cys		455	Ala	Leu	Asn	Asp	Leu 460	Cys	Ile	Lys	Val	13	92
465	ASII	TTP	чар	Leu	TTT Phe 470	Pne	ser	Pro	Ser	Glu 475	Asp	Asn	Phe	Thr	Asn 480	14	40
Asp	nea	ASII	цуѕ	485	GAA (Glu	lle	Thr	Ser 490	Asp	Thr	Asn	Ile	Glu 495	Ala	148	88
GCA Ala	GAA Glu	GAA Glu	Asn 500	ATT Ile	AGT Ser	TTA (Leu .	Asp	TTA Leu 505	ATA Ile	CAA Gln	CAA Gln	TAT Tyr	TAT Tyr 510	TTA Leu	ACC Thr	15:	36
TTT Phe	AAT Asn	TTT Phe 515	GAT Asp	AAT Asn	GAA (Pro (GAA . Glu . 520	AAT Asn	ATT Ile	TCA Ser	ATA Ile	GAA Glu 525	AAT Asn	CTT Leu	TCA Ser	158	34
AGT Ser	GAC Asp 530	ATT Ile	ATA Ile	GGC Gly	CAA '	TTA (Leu (535	GAA (Glu)	CTT Leu	ATG (Met	Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	163	32

545	550	GIU LEU AS	p Lys Tyr Th 555	TT ATG TTC CAT TAT ir Met Phe His Tyr 560	1680
CTT CGT GCT Leu Arg Ala	CAA GAA TTT Gln Glu Phe 565	GAA CAT GG Glu His Gl	T AAA TCT AG Y Lys Ser Ar 570	G ATT GCT TTA ACA g Ile Ala Leu Thr 575	1728
	580	589	i Pro Ser Arg	T GTT TAT ACA TTT g Val Tyr Thr Phe 590	1776
595	op 1/1 val	600	. Asn Lys Ala	T ACG GAG GCA GCT a Thr Glu Ala Ala 605	1824
610	dry trp var	615	val Tyr Asp 620		1872
625	630	INC ASP LYS	635	ATA ACT ATA ATT Ile Thr Ile Ile 640	1920
110 110 171	645	ta Leu Asn	650 Asn	ATG TTA TAT AAA Met Leu Tyr Lys 655	1968
6	660 ATA 1	665	Ser Gly Ala	GTT ATT CTG TTA Val Ile Leu Leu 670	2016
675	TO GIU IIE A	680	Val Leu Gly	ACT TTT GCA CTT Thr Phe Ala Leu 685	2064
GTA TCA TAT A Val Ser Tyr I 690	TE ATA ASII L	AG GTT CTA ys Val Leu 95	ACC GTT CAA Thr Val Gln 700	ACA ATA GAT AAT Thr Ile Asp Asn	2112
GCT TTA AGT A Ala Leu Ser L 705	AA AGA AAT G ys Arg Asn G 710	AA AAA TGG lu Lys Trp	GAT GAG GTC Asp Glu Val 715	TAT AAA TAT ATA Tyr Lys Tyr Ile 720	2160
TOTAL MOTE I	GG TTA GCA A rp Leu Ala L 725	s val Asn	ACA CAG ATT Thr Gln Ile 730	GAT CTA ATA AGA Asp Leu Ile Arg 735	2208
בין בין בין בין	AA GAA GCT T ys Glu Ala Le 40	TA GAA AAT (eu Glu Asn (745	CAA GCA GAA Gln Ala Glu	GCA ACA AAG GCT Ala Thr Lys Ala 750	2256
ATA ATA AAC TA Ile Ile Asn Ty 755	AT CAG TAT AM	T CAA TAT A IN Gln Tyr 1 760	Thr Glu Glu	GAG AAA AAT AAT Glu Lys Asn Asn 765	2304
ATT AAT TTT AA Ile Asn Phe As 770	AT ATT GAT GA sn Ile Asp As 77	b red ser s	CCG AAA CTT A Ser Lys Leu A 780	AAT GAG TCT ATA Asn Glu Ser Ile	2352
AAT AAA GCT AT Asn Lys Ala Me 785	G ATT AAT AT t Ile Asn Il 790	A AAT AAA 1 e Asn Lys p	TT TTG AAT (he Leu Asn (795	CAA TGC TCT GTT Gln Cys Ser Val 800	2400
TCA TAT TTA AT Ser Tyr Leu Me	G AAT TCT AT t Asn Ser Me 805	r ite bro i	AT GGT GTT A yr Gly Val I 10	AAA CGG TTA GAA Lys Arg Leu Glu 815	2448

GAT Asp	TTT Phe	GAT Asp	GCT Ala 820	SET	CTT	AAA Lys	GAT Asp	GCA Ala 825	TTA Leu	TTA Leu	AAG Lys	TAT	ATA Ile 830	TAT Tyr	GAT Asp	24	496
AAT Asn	AGA Arg	GGA Gly 835	ACT Thr	TTA Leu	ATT Ile	GGT Gly	CAA Gln 840	GTA Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 845	GAT Asp	AAA Lys	GTT Val	25	544
AAT Asn	AAT Asn 850	ACA Thr	CTT Leu	AGT Ser	ACA Thr	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	TCC Ser	AAA Lys	TAC Tyr	GTA Val	25	592
GAT Asp 865	AAT Asn	CAÁ Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCT Ser	AGG Arg 880	. 26	540
CCT Pro	GGA Gly	CCG Pro	GAG Glu	ACG Thr 885	CTC Leu	TGC Cys	GGG Gly	GCT Ala	GAG Glu 890	CTG Leu	GTG Val	GAT Asp	GCT Ala	CTT Leu 895	CAG Gln	26	88
TTC Phe	GTG Val	TGT Cys	GGA Gly 900	GAC Asp	AGG Arg	GGC Gly	TTT Phe	TAT Tyr 905	TTC Phe	AAC Asn	AAG Lys	CCC Pro	ACA Thr 910	GGG Gly	TAT Tyr	27	36
GGC Gly	TCC Ser	AGC Ser 915	AGT Ser	CGG Arg	AGG Arg	GCG Ala	CCT Pro 920	CAG Gln	ACA Thr	GGT Gly	ATC Ile	GTG Val 925	GAT Asp	GAG Glu	TGC Cys	27	84
TGC Cys	TTC Phe 930	CGG Arg	AGC Ser	TGT Cys	GAT Asp	CTA Leu 935	AGG Arg	AGG Arg	CTG Leu	Glu	ATG Met 940	TAT Tyr	TGC Cys	GCA Ala	CCC Pro	28	32
CTC Leu 945	AAG Lys	CCT Pro	GCC Ala	AAG Lys	TCA Ser 950	GCT Ala	GAA Glu	GCT Ala	TAG							28	62

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 954 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu
85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

- Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
- Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140
- Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155 160
- Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175
- Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190
- Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205
- Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220
- Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 230 235
- Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255
- Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270
- Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 275 280 285
- Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300
- Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 305 310 315 320
- Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 325 330 335
- Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 340 345 350
- Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 355 360 365
- Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr 370 380
- Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 385 390 395 400
- Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 405 410 415
- Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 425 430
- Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys
 435
 440
 445
- Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 450 455 460

- Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 465 470 475 480
- Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 485 490 495
- Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 500 505 510
- Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 515 520 525
- Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 530 540
- Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555 560
- Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 565 570 575
- Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 580 585 590
- Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 595 600 605
- Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 610 620
- Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 625 630 640
- Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys
- Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 660 665 670
- Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 675 680 685
- Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 690 695 700
- Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 715 720
- Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg
 725 730 735
- Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala
 740 745 750
- Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765
- Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
 770 780
- Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800
- Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815

288

As	p Ph	e As	p Ala 82	a Se	r Le	ı Lys	Asp	Ala 825	Leu	Leu	Lys	Туз	r Il		r Asp)
As	n Ar	g Gl	y Thi	r Lei	ı Ile	Gly	Gln 840	Val	Asp	Arg	Leu	Lys 845		p Ly	s Val	
As	n Ası 85	n Th:	r Leı	ı Sei	r Thr	Asp 855	Ile	Pro	Phe	Gln	Leu 860	Ser	Lys	s Ту	r Val	
As 86	p Ası 5	n Glr	n Arg	, Lei	Leu 870	Ser	Thr	Phe	Thr	Glu 875	Tyr	Ile	Lys	s Se	r Arg 880	
Pr	o Gly	Pro	o Glu	Thr 885	Leu	Cys	Gly	Ala	Glu 890	Leu	Val	Asp	Ala	Le: 899	ı Gln	
Pho	e Val	l Cys	900	Asp	Arg	Gly	Phe	Tyr 905	Phe	Asn	Lys	Pro	Thr 910		Tyr	
Gly	/ Ser	Ser 915	Ser	Arg	Arg	Ala	Pro 920	Gln	Thr	Gly	Ile	Val 925	Asp	Glu	Cys	
Cys	930	Arg	Ser	Cys	Asp	Leu 935	Arg	Arg	Leu	Glu	Met 940	Tyr	Cys	Ala	Pro	
Leu 945		Pro	Ala	Lys	Ser 950	Ala	Glu	Ala	•							
(2)	INF	ORMA	TION	FOR	SEQ	ID N	0: 1	5 :								
	(ii)	() () () (I) () () ()	A) LECUIATURE	ENGTH (PE: (RANI)POLC LE TY L: LME/K	H: 27 nucl DEDNE DGY: (PE:	TERI 24 b eic SS: line DNA CDS	ase acid doub ar (gen	pair le								
	(xi)	SEC	QUENC	E DE	SCRI	PTION	J: SI	EQ II	NO:	: 15:						
ATG Met 1	CAG Gln	TTC Phe	GTG Val	AAC Asn 5	AAG (CAG T Gln E	ne A	Asn T	yr I	AAG G Jys A	sp I	CCT (Val	AAC Asn 15	GGT Gly	48
GTT Val	GAC Asp	ATT Ile	GCC Ala 20	TAC Tyr	ATC I	AAA A Lys I	TT C	CA A Pro A 25	AC C	CC G la G	GC C	IAG A	ATG Met (CAG Gln	CCG Pro	96
GTG Val	AAG Lys	GCT Ala 35	TTC / Phe i	AAG . Lys :	ATT (CAT A His A	AC A sn L 40	AA A iys I	TC T le T	GG G	TT A al I	TT Cle F	CCG (GAA Glu	CGC Arg	144
GAT Asp	ACA Thr 50	TTT . Phe '	ACG A	AAC (Asn)	CCG C Pro C	BAA G Blu G 55	AA G lu G	GA G	AC T sp L	eu A	AC C sn P 60	CG C	CCG (CCG Pro	GAA Glu	192
GCA Ala 65	AAG (CAG (Gln '	GTG (Val I	CCA (GTT T Val S	CA T	AC T yr T	AC G	sp S	CA AG er Ti	CC T hr T	AT C	TG A	AGC . Ser '	ACA Thr	240

GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85

	-	•	10	0			. (1)	10	9 Me 5	с ъе	u Le	u Th	r Se 11	r Il O	C GTC e Val		5
•		11	5			, 41	120)	r in	L 110	e As _l	2 Th:	r Gl	u Le	G AAG u Lys		į
	130	ני				135	ASI	va.	r 116	e GII	1 Pro) Asp	Gl;	y Se	C TAC r Tyr	432	
145					150	neu.	. Val	TTE	: 116	155	Pro	Ser	Ala	a Ası	C ATT P Ile 160	480	
				165	Lys	361	PHE	GIÀ	170	GIu	Val	. Leu	Asr	1 Let 175		.528	
·- J		1	180	O.	261	1111	GIN	185	116	Arg	Phe	Ser	Pro 190	Ast	TTC Phe	576	
	-,	195		Jiu	Giu	261	200	GIU	vai	Asp	Thr	Asn 205	Pro	Leu	TTG Leu	624	
2	210	027	273	1110	ALG	215	wsb	Pro	АТА	val	220	Leu	Ala	His	GAG Glu	672	
225			7,10	Gly	230	Arg	Leu	lyr	GIY	235	Ala	Ile	Asn	Pro	240	720	
7129	vai	FIIC	Lys	245	AAC Asn	inr	Asn	Ala	Tyr 250	Tyr	Glu	Met	Ser	Gly 255	Leu	768	
020		501	260	Giu	GAA Glu	Leu	Arg	265	Phe	Gly	Gly	His	Asp 270	Ala	Lys	816	
1110	116	275	SEL	Leu	CAG Gln	GIU	280	GIu	Phe	Arg	Leu	Tyr 285	Tyr	Tyr	Asn	864	
2,3	290	пуъ	Asp	116		295	inr	Leu	Asn	Lys	Ala 300	Lys	Ser	Ile	Val	912	
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG '	TAT . Tyr i	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960	
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT A	ACA '	TCT (Ser (31A	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008	
цув	FIIE	Asp	340	Leu	TAC A	Lys i	met :	Leu '	Thr	Glu	Ile	Tyr	Thr 350	Glu	Asp	1056	
AAT ' Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT / Phe 1	Lys '	GTA (Val 1 360	CTT . Leu .	AAC . Asn .	AGA Arg	Lys '	ACA ' Thr ' 365	TAT Tyr	TTG Leu	AAT Asn	1104	

	370	273	710	Va.	. File	375	;	e As	in 11	Le Va	31 P:	ro L	ys V	al A	sn	_	•	1152
385	ATA Ile	IYL	Asp	GIY	390	AST	Let	u Ar	g As	n Th 39	r As	sn Le	eu A	la A	la	Asn 400		1200
FIIC	AAT Asn	GIY	GIII	405	Int	Glu	116	₽ As:	n As 41	n Me 0	t As	in Ph	ne Ti	1r L	ys 15	Leu		1248
270	AAT . Asn	1110	420	GIY	beu	Pne	GIU	42	e Ty	r Ly	s Le	u Le	u Cy 43	s Va 0	al	Arg		1296
GGG Gly		435	****	361	БУЗ	1111	440	sei	r Lei	u Ası	p Ly	s Gl 44	у Ту 5	r As	in	Lys		1344
	450	o r y	A. g	Cys	vsħ	455	мта	Let	ı Ası	n Asp	46	и Су	s Il	e Ly	's '	Val		1392
AAT A Asn A 465		· ·	чэħ	rea	470	Pne	Ser	Pro	Ser	475	ı Ası) Ası	n' Ph	e Th	r 1	Asn 180		1440
GAT (Asp I	JCQ F	.311 1	uys '	485	GIU	GIU	116	inr	490	Asp	Thi	Ası	ı Ile	e Gl:	u A 5	la		1488
GCA G Ala G	riu G	5	500	iie .	ser	Leu .	Asp	505	Ile	Gln	Gln	Tyr	Ty:	Le	ıΤ	hr		1536
TTT A Phe A	5	15	A GE	ASII (JIU !	Pro (520	Asn	He	Ser	Ile	Glu 525	Asn	Le.	ıS	er		1584
	30	16 1	16 0	sry (2111 1	35 535	31U	ren	Met	Pro	Asn 540	Ile	Glu	Arg	P	he		1632
CCT A Pro A 545	311 G .	. , .	ys L	,ys 1 5	50	oru r	zeu z	Asp	гÀг	Tyr 555	Thr	Met	Phe	His	T)	7r 50		1680
CTT CO	rg A	La G	5	65	ne G	itu B	iis (зтА	Lys 570	Ser	Arg	Ile	Ala	Leu 575	Tì	ır		1728
AAT TO Asn Se	si Va	5 E	90 90	IU A	.Ia L	eu L	eu A	Asn 585	Pro	Ser	Arg	Val	Tyr 590	Thr	Ph	ıe	:	1776
TTT TO Phe Se	r Se	T W	AC T	AT G yr V	TA A al L	λeri	AA C ys V 00	STT . /al .	AAT Asn	AAA Lys	GCT Ala	ACG Thr 605	GAG Glu	GCA Ala	GC Al	T a	1	L824
ATG TT Met Ph 61	ie ne	'A GC	SC TO Ly Tr	SG G	ar G	AA C lu G 15	AA T ln L	TA (GTA Val	Tyr /	GAT Asp 620	TTT Phe	ACC Thr	GAT Asp	GA G1	A u	1	1872
ACT AG Thr Se 625	C GA r Gl	A GT u Va	TA AC	er 11	CT AG hr Tl	CG G	AT A sp L	AA) ys :	Ile i	GCG (Ala / 635	GAT Asp	ATA Ile	ACT Thr	ATA Ile	AT 11 64	e	1	920

AT Il	T CCA	TA:	T ATA	GGA Gly 645	PEC	GCT Ala	TTA Leu	AA1 Asr	T ATA 116 650	: G1}	AA1 Asr	r ATO	TT/	A TA:	r AAA Lys		1968
GA: Asi	GAT Asp	TTT Phe	GTA Val 660	GIA	GCI Ala	TTA Leu	ATA Ile	TTT Phe 665	Ser	GGA Gly	GCT Ala	GTI Val	ATT Ile	Let	TTA Leu		2016
GA/ Glu	TT1 Phe	11e 675	Pro	GAG Glu	ATT	GCA Ala	ATA Ile 680	Pro	GTA Val	TTA Leu	GGT	ACT Thr 685	Phe	GCA Ala	CTT Leu		2064
GT# Val	TCA Ser 690	ıyı	ATT Ile	GCG Ala	AAT Asn	AAG Lys 695	GTT Val	CTA Leu	ACC Thr	GTT Val	CAA Gln 700	Thr	ATA	GAT Asp	AAT Asn		2112
705	Leu	ser	Lys	Arg	710	GIU	гуs	Trp	Asp	Glu 715	Val	Tyr	Lys	Tyr	720	. *	2160
val	1111	ASII	TGG Trp	725	AIA	rvs	Vai	Asn	730	Gln	Ile	Asp	Leu	Ile 735	Arg		2208
AAA Lys	AAA Lys	ATG Met	AAA Lys 740	GAA Glu	GCT Ala	TTA Leu	GAA Glu	AAT Asn 745	CAA Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 750	AAG Lys	GCT Ala		2256
ATA Ile	ATA Ile	AAC Asn 755	TAT Tyr	CAG Gln	TAT Tyr	AAT Asn	CAA Gln 760	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 765	AAA Lys	AAT Asn	AAT Asn		2304
116	770	Pne	AAT Asn	11e	Asp	775	Leu	Ser	Ser	Lys	Leu 780	Asn	Glu	Ser	Ile		2352
785	ьуs	Ата	ATG Met	11e	790	lie	Asn	Lys	Phe	Leu 795	Asn	Gln	Cys	Ser	Val 800		2400
Ser.	ıyı	Leu	ATG Met	805	Ser	Met	lie	Pro	810	Gly	Val	Lys	Arg	Leu 815	Glu		2448
Asp	Pne	Asp	GCT Ala 820	Ser	Leu	гЛг	Asp	Ala 825	Leu	Leu	Lys	Tyr	11e 830	Tyr	Asp		2496
Asn	Arg	835	ACT Thr	Leu	Ile	Gly	Gln 840	Val	Asp	Arg	Leu	Lys 845	Asp	Lys	Val		2544
Asn	850	Thr	CTT Leu	Ser	Thr	Asp 855	Ile	Pro	Phe	Gln	Leu 860	Ser	Lys	Tyr	Val		2592
GAT Asp 865	AAT Asn	CAA Gln	AGA Arg	Leu	TTA Leu 870	TCT . Ser	ACA Thr	TTT Phe	Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCT Ser	AGG Arg 880		2640
CCT Pro	CAA Gln	TCT Ser	AAA Lys	GTT Val 885	AAA Lys	AGA Arg	CAA Gln	Ile	TTT Phe 890	TCA Ser	GGC Gly	TAT Tyr	Gln	TCT Ser 895	GAT Asp		2688
ATT Ile	GAT Asp	ACA Thr	CAT . His . 900	AAT Asn	AGA Arg	ATT . Ile	Lys	GAT Asp 905	GAA Glu	TTA Leu	TGA *						2724

- (2) INFORMATION FOR SEQ ID NO: 16:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 908 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr
65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys
115 120 125

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 135 140

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 150 155

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175

Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 180 185 190

Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 195 200 205

Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 210 215 220

Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 225 230 235 240

Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 270

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 275 280 285

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu 330 Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 390 395 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 470 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 520 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 570 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 620 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 635 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys

- Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu
- Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 680 685
- Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn
- Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile
- Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 725 730
- Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 745
- Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn
- Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
- Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 795
- Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu
- Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 825
- Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val
- Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val
- Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg 870
- Pro Gln Ser Lys Val Lys Arg Gln Ile Phe Ser Gly Tyr Gln Ser Asp 890
- Ile Asp Thr His Asn Arg Ile Lys Asp Glu Leu
- (2) INFORMATION FOR SEQ ID NO: 17:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3042 base pairs

 - (B) TYPE: nucleic acid (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..3042
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

; <u> </u>	CCG Pro	CAG Gln	ATG Met 30	CAG Gln	GGC Gly	GCC Ala	AAC Asn	CCA Pro 25	Ile	Lys	ATC Ile	LAL	GCC Ala 20	Ile	Asp	Val
14	CGC Arg	GAA Glu	CCG Pro	ATT Ile 45	GTT Val	TGG	ATC Ile	AAA Lys	AAC Asn 40	CAT His	ATT Ile	AAG Lys	Pne	GCT Ala 35	AAG Lys	GTC Val
. 19	GAA Glu	CCG Pro	CCG Pro	CCG Pro	AAC Asn 60	TTG Leu	GAC Asp	GGA Gly	GAA Glu	GAA Glu 55	CCG Pro	AAC Asn	ACG Thr	TTT Phe	ACA Thr 50	GAT Asp
	Thr 80	Ser	Leu	Tyr	Thr	Ser 75	Asp	Tyr	Tyr	ser	70	CCA Pro	Val	GIH	Lys	65
28	Glu	Phe 95	Leu	Lys	Thr	Val	Gly 90	Lys	Leu	ıyr	ASN	GAT Asp 85	nys	GIU	ASII	чэр
33	Val	Ile	Ser 110	Thr	Leu	Leu	Met	Arg 105	GIY	rea	Asp	ACT Thr	100	IYL		wra
384	AAG Lys	TTG Leu	GAG Glu	ACG Thr 125	Asp	ATT Ile	ACC Thr	AGT Ser	GGC Gly 120	GGT Gly	TGG Trp	TTT Phe	CCA Pro	ATC Ile 115	GGA Gly	CGC Arg
432	TAC Tyr	AGC Ser	GGT Gly	GAC Asp	CCA Pro 140	Gln	ATC Ile	GTG Val	AAC Asn	ATT Ile 135	TGC Cys	AAC Asn	ACT Thr	GAC Asp	ATT Ile 130	GTT Val
480	ATT Ile 160	Asp	GCG (Ala ,	TCC (Ser /	CCC Pro	GGG Gly 155	Ile	ATC Ile	GTA Val	CTC Leu	AAC Asn 150	CTT Leu	GAA Glu	GAA Glu	TCT Ser	AGA Arg 145
528	ACG Thr	CTG / Leu 1	Asn :	TTG /	GTG '	GAA (Glu	CAC His 170	GIĀ	TTT Phe	AGC Ser	AAG Lys	TGC Cys 165	GAG Glu	TTT Phe	CAG Gln	ATC Ile
576	TTC Phe	GAC :	CCA (Pro 1	Ser 1	TTC / Phe s	CGT ' Arg '	ATT (TAC Tyr 185	Gin	ACT Thr	TCT Ser	GGC Gly	TAC Tyr 180	GIÅ	AAC Asn	CGT Arg
[^] 624	ITG Leu	TG T	CCG (Pro 1	AAC (Asn I 205	Thr A	GAT A	GTT (GAG Glu	CTG Leu 200	Ser	GAG Glu	GAG (Glu (TTC Phe	GGT Gly 195	Pne	ACG Thr
- 672	GAG Glu	CAC C	GCA (Ala H	CTG (Leu /	ACC (Thr I 220	Val :	GCG (Ala '	CCA (Pro	GAT Asp	ACT Thr 215	Ala	TTC (AAG Lys	GGC Gly	GCA Ala 210	GGT Gly
720	AAC Asn 240	ro A	AAC (Asn I	ATT A	GCG A	ATT (Ile / 235	Gly :	TAT (CTG Leu	CGT Arg	CAT His 230	GGT (Gly 1	GCC Ala	CAC His	ATC Ile	CTG Leu 225
768	TTA Leu	GT T ly 1 55	Ser C	ATG A Met S	GAG A Glu N	TAC (Tyr (TAC 1 Tyr 1 250	Ala '	AAC (Asn .	ACC . Thr .	AAC . Asn '	GTT A Val A 245	Lys	TTC . Phe	GTG Val	CGC Arg
816	Lys Lys	CG A	SAT C Asp #	lis P	GC C	GGT (TTC (Phe (ACG Thr 1 265	Arg '	CTG Leu	GAA (Glu)	GAG (Glu (TTC Phe 260	Ser .	GTA . Val	GAA Glu
864	NAC Asn	AC A	TAC T	TAC I Tyr I	eu 1	CGT (Arg I	TTC (Phe <i>l</i>	Glu :	AAC (Asn (280 -	Glu .	CAG (Gln (ITG (Leu (AGC Ser	GAC Asp :	ile .	rtt Phe

A L	AG 7 Ys I	TTT A Phe I	AAA (.ys <i>!</i>	ATA I qa	TT G	.ra 3	GT A er T 95	CA C	TG /	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	AT	T GTG e Val	912
3 (05	_			3	10	 1	YE M	et 1	ys .	Asn 315	Val	Phe	Lys	Gl	G AAA 1 Lys 320	
•				3.	25	aħ II	11 30	er G	3 TA 17	30	Phe	Ser	Val	Asp	Lys 335		1008
-2			3	40	-u + 1	r Dy	'S ME	34	eu 1 45	nr G	ilu	Ile	Tyr	Thr 350	Glu	GAT Asp	1056
		3 5	55	,	TT TT ne Ph	.c ny	36	0	zu A	sn A	rg i	Lys	Thr 365	Tyr	Leu	Asn	1104
,	37	0			A TT	37	5	e As	in, 11	le v	al E	880	Lys '	Val	Asn	Tyr	1152
38	5	1	- •••	p GI	A TT y Ph 39	0	ı Le	u Ar	g As	3:	nr A 95	sn I	Leu A	Ala .	Ala	Asn 400	1200
			, 01	40		. GI	1 110	e As	n As 41	n Me	et A	sn F	Phe I	hr i	Lys 415	Leu	1248
4			42	0	A TTO	• 6116	GI	425	e 1y	r Ly	rs L	eu L	eu C	30 30	/al	Arg	1296
•		439	5		C AAA C Lys		440) Sel	Lei	u As	b r	ys G 4	ly T 45	yr A	lsn	Lys	1344
	450)	••••	, cys	GAT Asp	455	Ala	Leu	ASI	1 As	P Le	eu Cy 50	ys I	le L	ys '	Val	1392
465		·	1,5	, DCG	TTT Phe 470	rne	261	PIO	Ser	47.	u As 5	ip As	sn Pl	ne T	hr A	Asn 180	1440
•			-, -	485	GAA Glu	GIU	116	Ing	490	Ası) Th	r As	sn Il	.e G 4	lu A 95	la	1488
			500	110	AGT Ser	Leu	Asp	505	rre	GIr	ı Gl	n Ty	r Ty 51	L Te	eu Ţ	hr	1536
		515	пор	ASII	GAA Glu	PIO	520	ASI	lle	Ser	· Ile	e Gl 52	u As 5	n Le	eu S	er	1584
	530			GLY	CAA Gln	535	GIU	Leu	met	Pro	54 (n Il.	e Gl	u Ar	g P	he	1632
CCT Pro 545	AAT Asn	GGA Gly	AAA Lys	AAG Lys	TAT Tyr 550	GAG Glu	TTA Leu	GAT Asp	AAA Lys	TAT Tyr 555	ACT Thi	T ATO	G TTO	C CA ≘ Hi	s T	AT Yr 50	1680

CT Le	T CC	T G	CT CA	AA GA Ln GI 56		TT GA ne Gl	A CA u Hi	T GG s Gl	T AA y Ly 57	s se	T AC	G AT	T GO	la L	TA 2 eu 1 75	ACA Th <u>r</u>	1728
			58	0		A TT. a Le	a ne,	58	5	o Se	r Ar	g Va	.1 Ty 59	T Th	ar E	Phe	1776
TT Ph	T TC e Se	T TC r Se 59		C TA p Ty	T GT r Va	A AAG l Lys	G AAA S Lys 600	· va.	r aa L Asi	r AA.	A GC s Al	T AC a Th 60	r Gl	G GC u Al	CA G	CT la	1824
	61	o			, .	A GAA 1 Glu 615	5	. rec	ı val	LTy	62	p Pho 0	e Th	r As	p G	lu	1872
625	5				630		. nsp	nys	, ite	635	ASI	o Ile	≘ Th:	r Il	e I 6	le 40	1920
		- 7		64	5	GCT Ala	. Deu	ASII	650	Gly	Asr Asr	ı Met	: Le	1 Ty 65	r L; 5	ys ·	1968
			660)	7120	TTA Leu	116	665	ser	GIA	Ala	. Val	. Ile	Le	u Le	eu	2016
		675	,	, 010		GCA Ala	680	Pro	vai	Leu	Gly	Thr 685	Phe	Ala	a Le	eu .	2064
	690	-1-		, ,,,,,	. ASI	AAG Lys 695	vai	reu	Inr	Vai	Gln 700	Thr	Ile	Asp) As	n	2112
705		501	Lly S	n. y	710	GAA Glu	Lys	Trp	Asp	715	Val	Tyr	Lys	Туг	72	e 0	2160
			1.5	725	AIG	AAG Lys	Vai	Asn	730	Gin	Ile	Asp	Leu	Ile 735	Ar	g	2208
AAA Lys	AAA Lys	ATG Met	AAA Lys 740	GAA Glu	GCT Ala	TTA Leu	GAA Glu	AAT Asn 745	CAA Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 750	AAG Lys	GC Al	T a	2256
ATA Ile	ATA Ile	AAC Asn 755	TAT Tyr	CAG Gln	TAT Tyr	AAT Asn	CAA Gln 760	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 765	AAA Lys	AAT Asn	AA' As:	r n	2304
ATT	AAT Asn 770	TTT Phe	AAT Asn	ATT Ile	GAT Asp	GAT Asp 775	TTA Leu	AGT Ser	TCG Ser	AAA Lys	CTT Leu 780	AAT Asn	GAG Glu	TCT Ser	ATZ Ile	A e	2352
AAT Asn 785	AAA Lys	GCT Ala	ATG Met	ATT Ile	AAT Asn 790	ATA Ile	AAT . Asn	AAA Lys	Phe	TTG Leu 795	AAT Asn	CAA Gln	TGC Cys	TCT Ser	GT: Val 800	L	2400
TCA Ser	TAT Tyr	TTA Leu	ATG Met	AAT Asn 805	TCT Ser	ATG Met	ATC (Pro '	TAT Tyr 810	GGT Gly	GTT Val	AAA Lys	CGG Arg	TTA Leu 815	GA/ Glu	A 1	2448
GAT Asp	TTT Phe	GAT Asp	GCT Ala 820	AGT Ser	CTT Leu	AAA Lys	Asp A	GCA ' Ala : B25	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 830	TAT Tyr	GA1 Asp		2496

AA: Asr	r AGA	A GG/ g Gly 835	A T111	r TTA	A ATT	r GGT e Gly	CAA Glr 840	ıval	A GAT	AGA Arg	TT:	A AAA 1 Lys 845	: As	r aa D Ly	A GTT s Val	2544
AA] Asr	AA7 Asr 850		A CTT	Γ AG1 1 Ser	ACA Thr	GAT Asp 855	Tre	CCI Pro	TTT Phe	CAG Gln	CTT Let 860	ı Ser	Lys	А ТАО в Тур	C GTA	2592
GAT Asp 865	L VOI	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	Tyr	T ATT	AAC Lys	TC. Ser	GGC Gly 880	2640
Бец	. ASII	361	PIO	885	AIA	AIA	HIS	Tyr	890	Gln	His	Asp	Glu	Ala 895		2688
GAC Asp	AAC Asn	AAA Lys	Phe 900	AAC Asn	AAA Lys	GAA Glu	CAA Gln	CAA Gln 905	AAC Asn	GCG Ala	TTC Phe	TAT Tyr	GAG Glu 910	ATC Ile	TTA Leu	2736
	Deu	915	ASII	ren	ASN	GIU	920	GIn	Arg	Asn	Ala	TTC Phe 925	Ile	Gln	Ser	2784
beu	930	Asp	Asp	Pro	ser	935	Ser	Ala	Asn	Leu _.	Leu 940	GCA Ala	Glu	Ala	Lys	2832
945	Leu	ASII	Asp	Ala	950	Ala	Pro	Lys	Val	Asp 955	Asn	AAA Lys	Phe	Asn	Lys 960	2880
Giu	GIII	GIII	ASN	965	Pne	Tyr	Glu	Ile	Leu 970	His	Leu	CCT Pro	Asn	Leu 975	Asn	2928
GAA Glu	GAA Glu	CAA Gln	CGA Arg 980	AAC Asn	GCC Ala	TTC . Phe	lle	CAA Gln 985	AGT Ser	TTA . Leu	AAA Lys	GAT Asp	GAC Asp 990	CCA Pro	AGC Ser	2976
CAA Gln	ser.	GCT Ala 995	AAC Asn	CTT Leu	TTA . Leu .	Ala (GAA (Glu / 1000	GCT Ala	AAA . Lys :	AAG (Lys)	Leu	AAT (Asn 1	GAT Asp	GCT Ala	CAG Gln	3024
Ala	CCG Pro 1010	AAA Lys	GTA Val .	GAC ' Asp	TAG											3042

(2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1014 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 135 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 470 475 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser 520 Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 570 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 615 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 665 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala-Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 710 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala

745

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 755 760 765

Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile
770 780

Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val 785 790 795 800

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815

Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp 820 825 830

Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 835 840 845

Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 850 855 860

Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Gly 865 870 875 880

Leu Asn Ser Pro Gly Ala Ala His Tyr Ala Gln His Asp Glu Ala Val 885 890 895

Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu 900 905 910

His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser 915 920 925

Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys 930 935 940

Lys Leu Asn Asp Ala Gln Ala Pro Lys Val Asp Asn Lys Phe Asn Lys 945 950 955 960

Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Prò Asn Leu Asn 965 970 975

Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser 980 985 990

Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln 995 1000 1005

Ala Pro Lys Val Asp * 1010

(2) INFORMATION FOR SEQ ID NO: 19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3509 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..3509
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

ME	1	.0 .	41 1	nr I	5 5	AT AA sn As	n Pn	e As	n Ty	r Ası	n As	p:Pr 	o Il	e As	sp . 15	Asn		48
AA As	T AP	AT A' sn I!	Le I	TT AT le Me 20	TG AT	G GA	G CC	T CC. o Pro 2	o Phe	r GCC e Ala	G AGA	A GG g Gl	y Th	G G(x G] 0	GG /	AGA Arg		96
TA Ty	T TA r Ty	r r	NA GO vs Ai 15	CT TT la Ph	TT AA ne Ly	A ATO	C ACA	r Ası	CG1	T ATT	TGC Trp	G ATA	e Il	A CC e Pr	G (GAA Glu	1	44
AG. Ar	ATA Ty S	r ir	T T	TT GG	SA TA .y Ty	T AAA T Lys	Pro	GAC Glu	G GAT	TTT Phe	AAT Asn 60	Lys	A AG	T TC r Se	C C	GT Sly	1	92
AT Ile 6	e Pn	T AA e As	T AC	A GA	T GT p Va. 7	T TGT l Cys 0	GAA Glu	TAT	TAT	GAT Asp 75	Pro	GAT Asp	TAC Ty	TT Le	A A	AT sn 80	24	40
ACT Thi	AA 1 Asi	T GA n As	T AA p Ly	s гх	G AA' s Asi 5	T ATA	TTT Phe	TTA Leu	CAA Gln 90	Thr	ATG Met	ATC Ile	AAC Lys	TT.	u P	TT he	28	38
AAT Asr	AGA	A AT	C AA e Ly 10	s Se	A AAA r Lys	A CCA S Pro	TTG Leu	GGT Gly 105	GAA Glu	AAG Lys	TTA Leu	TTA Leu	GAG Glu 110	Met	G A	TT le	. 33	16
ATA Ile	AAT Asr	GG Gl; 115	\ TT	A CC' e Pro	TA1	CTT Leu	GGA Gly 120	GAT Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	CTC Leu	GA/ Glu	A G.	AG lu	38	4
TTT Phe	AAC Asn 130	Thi	A AA	C ATT	r GCT e Ala	AGT Ser 135	GTA Val	ACT Thr	GTT Val	AAT Asn	AAA Lys 140	TTA Leu	ATC Ile	AG1 Ser	A.	AT sn	43	2
CCA Pro 145	GIY	GA# Glu	GT(LGIU	G CGA Arg 150	AAA Lys	AAA Lys	GGT Gly	ATT Ile	TTC Phe 155	GCA Ala	AAT Asn	TTA Leu	ATA Ile	A1	.e	48	o
TTT Phe	GGA Gly	CCT Pro	GGG Gly	CCA Pro 165	Val	TTA Leu	AAT Asn	GAA Glu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	Asp	ATA Ile 175	Gl	T Y	528	8
ATA Ile	CAA Gln	AAT Asn	CAT His	Phe	GCA Ala	TCA Ser	AGG Arg	GAA Glu 185	GGC Gly	TTC Phe	GGG Gly	GGT Gly	ATA Ile 190	ATG Met	CA Gl	A n	576	5
ATG Met	AAG Lys	TTT Phe 195	Cys	CCA Pro	GAA Glu	TAT Tyr	GTA Val 200	AGC Ser	GTA Val	TTT . Phe .	Asn .	AAT Asn 205	GTT Val	CAA Gln	GA Gl	A u	624	Į
AAC Asn	AAA Lys 210	GGC Gly	GCA Ala	AGT Ser	ATA Ile	TTT Phe 215	AAT Asn	AGA Arg	CGT (Gly 1	TAT 1 Tyr 1 220	TTT Phe	TCA Ser	GAT Asp	CC.	A O	672	
GCC Ala 225	TTG Leu	ATA Ile	TTA Leu	ATG Met	CAT His 230	GAA Glu	CTT . Leu	ATA (His Y	GTT : Val 1 235	TTA (Leu I	CAT (GGA Gly	TTA Leu	TA: Ty: 240	r	720	
GGC	ATT Ile	AAA Lys	GTA Val	GAT Asp 245	GAT Asp	TTA (CCA /	Ile '	GTA (Val 1 250	CCA # Pro #	AAT (Asn (GAA 1 Glu 1	Lys	AAA Lys 255	TTT Phe	r =	768	
TTT Phe	ATG Met	CAA Gln	TCT Ser 260	ACA Thr	GAT Asp	GCT A	Ile (CAG (Gln # 265	GCA (GAA G	BAA C	Leu 1	TAT I	ACA Thr	TT1 Phe	Γ :	816	

GGA Gly	GGA Gly	CAA Gln 275	Asp	CCC Pro	AGC Ser	ATC Ile	ATA Ile 280	Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	Lys	AGT Ser	ATC		864
TAT Tyr	GAT Asp 290	rys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	Asp	AGA Arg	CTT Leu	AAC Asn		912
AAG Lys 305	vai	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320		960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly		1008
AAA Lys	TAT	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu		1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys		1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT	TTT	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys		1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	;	1200
TCT	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	:	1248
AAT Asn	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	AGC Ser 425	Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	:	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	j	L344
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	. 1	L392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1	1440
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1	.488
TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	1	.536
GAT Asp	TTT Phe	AAT Asn 515	GTA Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	TAT Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys	1	.584
AAA Lys	ATT Ile 530	TTT Phe	ACA Thr	GAT Asp	GAA Glu	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC Tyr	TCT Ser	CAG Gln	1	.632

AAT GCT TTT GAG ATT GCA GGA GCC AGT ATT CTA GAA TTT ATA CCAA ASN Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pho Gaa CTT TTA ATA CCT GTA GTT GGA GCC TTT TTA TTA GAA TCA TAT ACGAA CTT TTA ATA CCT GTA GTT GGA GCC TTT TTA TTA GAA TCA TAT ACGAA AAT AAA AAT AAA ATT ATT AAA ACA ATA GAT AAT GCT TTA ACT ACGAA ASN ASN Lys Asn Lys Ile Ile Lys Thr Ile Asn Asn Asn Ala Leu Thr Ly 685 AGA AAT GAA AAA TGG AGT GAT ATG TAC GGA TTA ATA GTA GCC CAA TCA GAT ASN ASN Glu Lys Trp Ser Asn Met Tyr Gly Leu Ile Val Ala Gln Tr 690 CTC TCA ACA GTT AAT ACT CAA TTT TAT ACA ATA AAA AGA ATG GCT TTA ACT ACGAC GCC CAA TCA TTA TTA CAA ACA ATA AAA AGA ATA AAA ATG GCC CAA TCA TTA TTA CAA ACA ATA AAA AGA GGA ATG TACC GAA TTA ATA GTA GCG CAA TCA GCC CAA TCA ACA GCA GCA TTA ATA GTA GCG CAA TCA TTA TACCAA TTA TACCAA TTA TACCAA TTA TACCAA TTA TACCAA TTA AAA GAG GGA ATG TACCAA TTA AAA ACA ATA AAA GAG GGA ATG TACCAA TTA TACCAA TACCAA TTA TACCAA TTA TACCAA TTA TACCAA TTA TACCAA TTA TACCAA TACAA TTA TACCAA TACCAA TTA TACCAA TACCAA TTA TACCAA TACAA TACAA TTA TACCAA TACAA TACAA TACAA TTA TACCAA TACAA TACAA TACAA TTA TACCAA TACAA TACAA TACAA TACAA TACAA TACAA TACAA TACAA TACAA TACAAA		ACA Thr 545	Phe	r cc e Pr	T CT O Le	ra G/ eu As	AT AT sp Il 55	e Ar	A GA: g Asi	r AT	A AG e Se	T TI r Le 55	u Th	CA TO	T TO	CA Ti er Pi	TT GA ne As 56	p .	168	0
TYP ILE Lys The Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala G 590 TGG GTG AAA CAG ATA GTA AAT GAT TTT GTA ATC GAA GCT AAT AAA AA TTP Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys S 595 AAT ACT ATG GAT AAA ATT GCA GAT ATA TCT CTA ATT GTT CCT TAT A ASN THY Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr I 610 GGA TTA GCT TTA AAT GTA GGA AAT GAA ACA GCT AAA GGA AAT TTT G GLY Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe G 635 AAT GCT TTT GAG ATT GCA GGA GCC AGT ATT CTA CTA GAA TTT ATA CC Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Phe 655 GAA CTT TTA ATA CCT GTA GTT GGA GCC TTT TTA TTA GAA TCA TAT ATA CLU Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr II 676 GAC AAT AAA AAT AAA ATT ATT AAA ACA ATA GAT AAT GCT TTA ACT AA ASP ASN Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Ly 675 AGA AAT GAA AAA TGG AGT GAT GT TAC GGA TTA ATA GAA GGA ATG TACA GAA ASP ASN Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Tr 700 CTC TCA ACA GTT AAT ACT CAA TTT TAT ACA ATA AAA GAG GGA ATG TACA GAG GCT TTA AATA GAT CAA TACA GAT AATA GAT AATA GAT CAA TACA GAT CAA GAT AATA GAT AATA TACA CAA TTA AATA CAA CAA TTA AATA CAA ATA AAA ATA AATA TACT CAA TTT TAT ACA ATA AAA GAG GGA ATG TACA GAG GCT TTA AATA GAT CAA CAA GAT AATA GAT CAA CAA GAT AATA GAT CAA CAA TACA CAA CAA CAA CAA CAA CAA		GAT Asp	GCA Ala	A TT a Le	A TI u Le	u Pr	ie Se	T AA r As	C AA/ n Lys	A GT	l Ty:	r Se	A TI	T TI e Ph	T TO	er Me	t As	T P	172	8
TTP Val Lys Gin lie Val Asn Asp Phe Val île Glu Ala Asn Lys S 605 AAT ACT ATG GAT AAA ATT GCA GAT ATA TCT CTA ATT GTT CCT TAT A Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr I 610 GGA TTA GCT TTA AAT GTA GGA AAT GAA ACA GCT AAA GGA AAT TTT GGY Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe G 635 AAT GCT TTG GAG ATT GCA GGA GCC AGT ATT CTA CTA GAA TTT ATA CA Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Ala Gly Ala Phe Leu Leu Glu Phe Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr II 6665 GAC AAT AAA AAT AAA ATT ATT AAA ACA ATA GAT AAT GCT TTA ACT AA Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Ly 670 AGA AAT GAA AAA TGG AGT GAT ATG TAC GGA TTA ATA GTA GCG CAA TCA ATA GAT AAT GAA AAT GAA AAT GAA AAT GAT ATA GTA G		TAT Fyr	ATT Ile	Γ AA ≥ Ly	s Th	r Al	T AA a As	T AA n Lys	A GTO	. Va	l Gl	A GC u Al	A GG a Gl	A TT y Le	u Ph	e Al	A GG a Gl	T Y	1776	5
GGA TTA GCT TTA AAT GCA GGA GCC AGT ATT CTA GAA TTA ACA ATA AAA AAT AAA AAT AAA AAT CAA GGA AAT GAA ACA GCT TTA AAT GCT GTA GAA ATT TTA GGA CTA CTA AAA GGA AAT TTT GGAA CTT TTA AAT CCT GTA GTA GGA GCC AGT ATT CTA CTA GAA TTA AAT AAT AAT AAT AAT AAT AAT A	•	rgg Crp	GTC Val	L Ly:	s GI	G AT n Il	A GT	A AAT 1 Asi	ı Asp	Phe	GTA Val	A AT	C GA e Gl	u Al	a As	T AA n Ly	A AG s Se	C:	1824	ı
GLY Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe G 635 630 630 630 630 635 630 635 630 635 636 635 636 635 636 635 636 635 636 635 636 635 636 635 636 635 636 635 636 635 636 635 636 636	Į	AAT Asn	Inr	Met	G GA E As	T AA p Ly	A AT s Ile	e Ala	Asp	ATA Ile	TCI Ser	CTI Let	ı Il	e Va	T CC l Pr	Т ТА о Ту	T AT	A e	1872	?
ASS ALA PRO GLU ILE ALA GLY ALA SET ILE LEU LEU GLU PRO 655 GAA CTT TTA ATA CCT GTA GTT GGA GCC TTT TTA TTA GAA TCA TAT AT GAA CTA TAT AT GAA CTA TAT AT GAA CTA TAT AT	C	IA	TTA Leu	GCT Ala	TT.	A AA u As:	n Val	l Gly	AAT Asn	GAA Glu	ACA Thr	Ala	Lys	A GGZ S Gly	A AA' Y Asi	T TT	T GAM e Glu 640	ı	1920)
GAC AAT AAA AAT AAA ATT ATT ATA ACA ATA AAA GAG GGA ATG TAGE Ser Tyr III AAT AAT AAT AAT AAT AAT AAA AAT TAT AAA ACA ATA GAT AAA AAT AAA ATT ATT AAA ACA ATA GAT AAT ACA ATA GAT ACA ATA ACA GCG CAA TO ACA ACA ACA ACA ACA ACA ACA ACA ACA AC	A	AT sn	GCT Ala	TT7 Phe	GA(ı Ilı	e Ala	GGA Gly	GCC Ala	AGT Ser	Ile	Leu	CT/ Lei	A GA/ 1 Gli	A TT	e Ile	e Pro	A	1968	
ASP ASN Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Ly AGA AAT GAA AAA TGG AGT GAT ATG TAC GGA TTA ATA GAG GCG CAA TG Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Tr 690 CTC TCA ACA GTT AAT ACT CAA TTT TAT ACA ATA AAA GAG GGA ATG Leu Ser Thr Val Asn Thr Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Ty 705 AAG GCT TTA AAT TAT CAA GCA CAA GCA TTG GAA GAA ATA ATA AAA TA Lys Ala Leu Asn Tyr Gln Ala Gln Ala Gln Glu Glu Ile Ile Lys Ty 735 AGA TAT AAT ATA TAT TCT GAA AAA GAA AAG TCA AAT ATT AAC ATC GA Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Gly Ser Asn Ile Asn 1le As 740 TTT AAT GAT ATA AAT TTT ATA AAT CTT AAT GAG GGT ATT AAC CAA GCT AT Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Il 755 GAT AAT ATA AAT AAT TTT ATA AAT GGA TGT TCT GTA TCA TAT TTA ATA Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Me 770 AAAA AAA ATG ATT CCA TTA GCT GTA GAA AAA TTA CTA GAC TTT GAT AAC Lys Met Ile Pro Leu Ala Val Glu Lys Leu Asp Phe Asp Asp 785 ACCT CTC AAA AAA AAA AAT TTG TTA AAT TAT ATA GAT GAA AAT AAA TTA TAT	G	AA lu	CTT Leu	TTA Leu	Ile	e Pro	GTA Val	GTT Val	GGA	Ala	TTT Phe	TTA Leu	TTA Leu	GAA Glu	ı Sei	Ty	T ATT	:	2016	
Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Tr 690 CTC TCA ACA GTT AAT ACT CAA TTT TAT ACA ATA AAA GAG GGA ATG TA Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Ty 715 AAG GCT TTA AAT TAT CAA GCA CAA GCA TTG GAA GAA ATA ATA AAA TA Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Ty 725 AGA TAT AAT ATA TAT TCT GAA AAA GAA AAG TCA AAT ATT AAC ATC GA Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asn 740 TTT AAT GAT ATA AAT TCT AAA CTT AAT GAG GGT ATT AAC CAA GCT AT Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Il 765 GAT AAT ATA AAT AAT TTT ATA AAT GGA TGT TCT GTA TCA TAT TTA ATA Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Me 770 AAA AAA ATG ATT CCA TTA GCT GTA GAA AAA TTA CTA GAC TTT GAT AAC Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn 790 ACT CTC AAA AAA AAA AAT TTG TTA AAT TAT ATA GAT GAA AAT AAA TTA TAT	G A	AC sp	AAT Asn	Lys	Asr	AAA Lys	ATT Ile	ATT Ile	Lys	ACA Thr	ATA Ile	GAT Asp	AAT Asn	Ala	Leu	ACT Thi	Lys		,2064	
Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Ty 705 AAG GCT TTA AAT TAT CAA GCA CAA GCA TTG GAA GAA ATA ATA AAA TA Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Ty 735 AGA TAT AAT AAT ATA TAT TCT GAA AAA GAA AAG TCA AAT ATT AAC ATC GA Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile As 740 TTT AAT GAT ATA AAT TCT AAA CTT AAT GAG GGT ATT AAC CAA GCT AT Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Il 755 GAT AAT ATA AAT AAT TTT ATA AAT GGA TGT TCT GTA TCA TAT TTA AT Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Me 770 AAAA AAA ATG ATT CCA TTA GCT GTA GAA AAA TTA CTA GAC TTT GAT AAC ASP Asn Ser Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asi 785 ACT CTC AAA AAA AAA AAT TTG TTA AAT TAT ATA GAT GAA AAT AAA TTA TAT	A	rg	Asn	GAA Glu	AAA Lys	TGC Trp	AGT Ser	Asp	ATG Met	TAC Tyr	GGA Gly	TTA Leu	Ile	Val	GCG Ala	CAA Gln	TGG Trp	;	2112	
AGA TAT AAT ATA AAT TCT AAA CTT AAT GGA TGT TCT GTA TAT TTA ATT AAC AAT AAT AAT AAT AAT A	L	eu .	TCA Ser	ACA Thr	GTT Val	AAT Asn	Thr	CAA Gln	TTT Phe	TAT Tyr	ACA Thr	Ile	AAA Lys	GAG Glu	GGA Gly	ATG Met	TAT Tyr 720		2160	
Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asn 740 TTT AAT GAT ATA AAT TCT AAA CTT AAT GAG GGT ATT AAC CAA GCT AT Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Il 755 GAT AAT ATA AAT AAT TTT ATA AAT GGA TGT TCT GTA TCA TAT TTA ATA ASP Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Me 770 AAA AAA ATG ATT CCA TTA GCT GTA GAA AAA TTA CTA GAC TTT GAT AAC Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Asp Phe Asp Asn 790 ACT CTC AAA AAA AAA AAT TTG TTA AAT TAT ATA GAT GAA AAT AAA TTA TAT	A/ Ly	AG (GCT Ala	TTA Leu	AAT Asn	Tyr	Gln	GCA Ala	CAA Gln	GCA Ala	Leu	GAA Glu	GAA Glu	ATA Ile	ATA Ile	Lys	TAC Tyr		2208	
Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Il 765 GAT AAT ATA AAT AAT TTT ATA AAT GGA TGT TCT GTA TCA TAT TTA AT Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Me 770 AAA AAA ATG ATT CCA TTA GCT GTA GAA AAA TTA CTA GAC TTT GAT AA' Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Asp Phe Asp Asp 785 ACT CTC AAA AAA AAA AAT TTG TTA AAT TAT ATA GAT GAA AAT AAA TTA TAT	A(A)	SA '	TAT Tyr	AAT Asn	Ile	TAT Tyr	TCT Ser	GAA Glu	AAA Lys	Glu	AAG Lys	TCA Ser	AAT Asn	ATT Ile	Asn	ATC Ile	GAT Asp		2256	
ASP ASN Ile ASN ASN Phe Ile ASN Gly Cys Ser Val Ser Tyr Leu Me 770 775 780 AAA AAA ATG ATT CCA TTA GCT GTA GAA AAA TTA CTA GAC TTT GAT AA Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Ass 785 790 795 800 ACT CTC AAA AAA AAT TTG TTA AAT TAT ATA GAT GAA AAT AAA TTA TA	T]	T /	AAT Asn	Asp	ATA Ile	AAT Asn	TCT Ser	AAA Lys	Leu	AAT Asn	GAG Glu	GGT Gly	ATT Ile	Asn	CAA Gln	GCT Ala	ATA Ile		2304	
Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asi 785 790 795 800 ACT CTC AAA AAA AAT TTG TTA AAT TAT ATA GAT GAA AAT AAA TTA TA	GA As	p A	Asn	ATA Ile	AAT Asn	AAT Asn	TTT Phe	Ile	AAT (Asn (GGA Gly	TGT Cys	TCT Ser	Val	TCA Ser	TAT Tyr	TTA Leu	ATG Met		2352	
ACT CTC AAA AAA AAT TTG TTA AAT TAT ATA GAT GAA AAT AAA TTA TAT	Ly	s I	YAA .	ATG Met	ATT Ile	CCA Pro	Leu	GCT Ala	GTA (Val (GAA Glu	Lys	Leu	CTA Leu	GAC Asp	TTT Phe	GAT Asp	AAT Asn 800		2400	
Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Ty 805 810 815	AC Th	T C	TC . Leu :	AAA Lys	AAA Lys	Asn	TTG Leu	TTA Leu	AAT 1 Asn 1	Гуr	Ile	GAT Asp	GAA Glu	AAT Asn	AAA Lys	Leu	TAT Tyr		2448	

	TTC Leu	ATT	GG/	A AG7 / Se1 820	. Alc	A GAA	TAT Tyr	GAA Glu	A AA/ Lys 825	Ser	AAA Lys	A GT	A AAT L Asr	AAA Lys 830	Tyr	TTG Leu	:	2496
	AAA Lys	ACC	11e 835	Mer	CCC Pro	TTT Phe	GAT Asp	CTI Leu 840	Ser	ATA	TAT	Thi	C AAT Asr 845	. Asp	ACA Thr	ATA Ile		2544
	CTA Leu	ATA Ile 850	GIU	ATG Met	TTT Phe	AAT Asn	AAA Lys 855	TAT Tyr	' AAT Asn	AGC Ser	GAA Glu	ATT	: Leu	AAT Asn	'AAT Asn	ATT	•	2592
	ATC Ile 865	Leu	AAT Asn	TTA Leu	AGA Arg	TAT Tyr 870	AAG Lys	GAT Asp	AAT Asn	AAT Asņ	TTA Leu 875	ATA Ile	GAT Asp	TTA Leu	TCA Ser	GGA Gly 880		2640
	TAT Tyr	GGG Gly	GCA Ala	AAG Lys	GTA Val 885	GAG Glu	GTA Val	TAT Tyr	GAT Asp	GGA Gly 890	GTC Val	GAG Glu	CTT Leu	AAT Asn	GAT Asp 895	AAA Lys		2688
	AAT Asn	CAA Gln	TTT Phe	AAA Lys 900	TTA Leu	ACT Thr	AGT Ser	TCA Ser	GCA Ala 905	AAT Asn	AGT Ser	AAG Lys	ATT	AGA Arg 910	GTG Val	ACT Thr		2736
	GIN	Asn	915	ASN	TIE	TIE	Phe	920	Ser	Val	Phe	Lėu	GAT Asp 925	Phe	Ser	Val		2784
	ser	930	ırp	iie	Arg	ile	935	Lys	Tyr	Lys	Asn	Asp 940	GGT Gly	Ile	Gln	Asn		2832
	945	11e	HIS	Asn	GIu	950	Thr	Ile	Ile	Asn	Cys 955	Met	AAA Lys	Asn	Asn	Ser 960		2880
	GIÀ	тр	Lys	TIE	965	lie	Arg	Gly	Asn	Arg 970	Ile	Ile	TGG Trp	Thr	Leu 975	Ile		2928
	GAT Asp	ATA Ile	AAT Asn	GGA Gly 980	AAA Lys	ACC Thr	AAA Lys	TCG Ser	GTA Val 985	TTT Phe	TTT	GAA Glu	TAT Tyr	AAC Asn 990	ATA Ile	AGA Arg		2976
(Glu	Asp	995	Ser	Glu	Tyr	Ile	Asn 1000	Arg	Trp	Phe	Phe	GTA Val 1005	Thr	Ile	Thr		3024
i	Asn	AAT Asn 1010	Leu	AAT Asn	AAC Asn	GCT Ala	AAA Lys 1015	Ile	TAT Tyr	ATT Ile	AAT Asn	GGT Gly 1020	AAG Lys)	CTA Leu	GAA Glu	TCA Ser		3072
- 1	AAT Asn 1025	Thr	GAT Asp	ATT Ile	Lys	GAT Asp 1030	Ile	AGA Arg	GAA Glu	GTT Val	ATT Ile 1035	Ala	AAT Asn	GGT Gly	GAA Glu	ATA Ile 1040		3120
]	ATA '	TTT Phe	AAA Lys	TTA Leu	GAT Asp 1045	Gly	GAT Asp	ATA Ile	GAT Asp	AGA Arg 1050	Thr	CAA Gln	TTT Phe	ATT Ile	TGG Trp 1055	Met		3168
I	AAA '	TAT Tyr	TTC Phe	AGT Ser 1060	Ile	TTT Phe	AAT Asn	Thr	GAA Glu 1065	TTA Leu	AGT Ser	CAA Gln	TCA Ser	AAT Asn 1070	ATT Ile	GAA Glu		3216
Ċ	GAA /	AGA Arg	TAT Tyr 1075	Lys	ATT Ile	CAA Gln	Ser	TAT Tyr 1080	Ser	GAA Glu	TAT Tyr	TTA Leu	AAA Lys 1085	Asp	TTT Phe	TGG Trp		3264

GGA Gly	AAT Asn 109	Pro	TTA Leu	ATG Met	TAC	AAT Asn 109	Lys	GAA Glu	TAT Tyr	TAT Tyr	ATG Met 110	TTT Phe 0	AAT Asn	GCG Ala	GGG Gly	3312
AAT Asn 110	Lys	AAT Asn	TCA Ser	TAT Tyr	ATT Ile 1110	Lys	CTA Leu	AAG Lys	AAA Lys	GAT Asp 111	Ser	CCT Pro	GTA Val	GGT Gly	GAA Glu 1120	3360
ATT Ile	TTA Leu	ACA Thr	CGT Arg	AGC Ser 1125	Lys	TAT Tyr	AAT Asn	CAA Gln	AAT Asn 1130	Ser	AAA Lys	TAT Tyr	ATA Ile	AAT Asn 1139	Tyr	3408
AGA Arg	GAT Asp	TTA Leu	TAT Tyr 1140	Ile	GGA Gly	GAA Glu	AAA Lys	TTT Phe 1149	Ile	ATA Ile	AGA Arg	AGA Arg	AAG Lys 1150	Ser	AAT Asn	3456
TCT Ser	CAA Gln	TCT Ser 1155	Ile	AAT Asn	GAT Asp	GAT Asp	ATA Ile 1160	Val	AGA Arg	AAA Lys	GAA Glu	GAT Asp 1165	Tyr	ATA Ile	TAT Tyr	3504
CTA Leu	GA															3509

(2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1169 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 10 15

Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg

Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35 40

Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 60

Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80

Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe
85 90 95

Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110

Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125

Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly
165 170 175

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 200 Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 230 Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 265 Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 295 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 330 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 345 Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 395 Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 405 Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 435 Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 455 Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys

Lys lle Phe Thr Asp Glu Asn Thr lle Phe Gln Tyr Leu Tyr Ser Gln 530 540

Thr Phe Pro Leu Asp lle Arg Asp lle Ser Leu Thr Ser Car De Car

Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 545 550 555 560

Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp 565 570 575

Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly 580 585 590

Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser 595 600 605

Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile 610 620

Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu 625 630 635 640

Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 645 650 655

Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile 660 665 670

Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys 675 680 685

Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp 690 695 700

Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr 705 710 715 720

Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr 725 730 735

Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp 740 745 750

Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile
755 760 765

Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met 770 780

Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn 785 790 795 800

Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr 805 810 815

Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu 820 825 830

Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile 835 840 845

Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Asn Asn Ile 850 855 860

Ile Leu Asn Leu Arg Tyr Lys Asp Asn Asn Leu Ile Asp Leu Ser Gly 875 880

Tyr Gly Ala Lys Val Glu Val Tyr Asp Gly Val Glu Leu Asn Asp Lys 885 890 895

Asn Gln Phe Lys Leu Thr Ser Ser Ala Asn Ser Lys Ile Arg Val Thr 900 905 910

Gln Asn Gln Asn Ile Ile Phe Asn Ser Val Phe Leu Asp Phe Ser Val 915 920 925

Ser Phe Trp Ile Arg Ile Pro Lys Tyr Lys Asn Asp Gly Ile Gln Asn 930 935 940

Tyr Ile His Asn Glu Tyr Thr Ile Ile Asn Cys Met Lys Asn Asn Ser 945 950 955 960

Gly Trp Lys Ile Ser Ile Arg Gly Asn Arg Ile Ile Trp Thr Leu Ile 965 970 975

Asp Ile Asn Gly Lys Thr Lys Ser Val Phe Phe Glu Tyr Asn Ile Arg 980 985 990

Glu Asp Ile Ser Glu Tyr Ile Asn Arg Trp Phe Phe Val Thr Ile Thr 995 1000 1005

Asn Asn Leu Asn Asn Ala Lys Ile Tyr Ile Asn Gly Lys Leu Glu Ser 1010 1015 1020

Asn Thr Asp Ile Lys Asp Ile Arg Glu Val Ile Ala Asn Gly Glu Ile 1025 1030 1035 1040

Ile Phe Lys Leu Asp Gly Asp Ile Asp Arg Thr Gln Phe Ile Trp Met 1045 1050 1055

Lys Tyr Phe Ser Ile Phe Asn Thr Glu Leu Ser Gln Ser Asn Ile Glu 1060 1065 1070

Glu Arg Tyr Lys Ile Gln Ser Tyr Ser Glu Tyr Leu Lys Asp Phe Trp 1075 1080 1085

Gly Asn Pro Leu Met Tyr Asn Lys Glu Tyr Tyr Met Phe Asn Ala Gly 1090 1095 1100

Asn Lys Asn Ser Tyr Ile Lys Leu Lys Lys Asp Ser Pro Val Gly Glu 1105 1110 1115 1120

Ile Leu Thr Arg Ser Lys Tyr Asn Gln Asn Ser Lys Tyr Ile Asn Tyr 1125 1130 1135

Arg Asp Leu Tyr Ile Gly Glu Lys Phe Ile Ile Arg Arg Lys Ser Asn 1140 1145 1150

Ser Gln Ser Ile Asn Asp Asp Ile Val Arg Lys Glu Asp Tyr Ile Tyr 1155 1160 1165

Leu

(2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2574 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..2574

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

			-		, -												
Me	G CC t Pr 1	A G1 o Va	TT AC	CA AT	TA AA .e As 5	T AA' n Asi	r TT n Ph	T AA e As	T TA n Ty 1	r Ası	T GA	T CC	T AT	e As	AT AAT Sp Asn .5		48
ASI	n As	11 11	.e 11	20 20	с ме	c GI	ı Pr	2 Pro	o Pho 5	e Ala	a Ar	g Gl	7 Th:	r Gl	G AGA y Arg		96
171	LIY	3 3	5 5	a Pn	е гу	s 116	4 (r Ası) Arg	J Ile	Tr	9 Ile 49	i Ile	e Pr	G GAA o Glu	1	44
ALC	5 1 y . 5 ()	r Pn	e GI	y 1y:	55 55	Pro	GIU	ı Asp	Phe	Asr 60	l Lys	Ser	Se	C GGT r Gly	19	92
65	FILE	E ASI	n Ar	g Asi	7 (Cys	GIU	ı Tyr	Tyr	75	Pro	Asp	Tyr	Lei	TAA A nzA ı 08	24	10
THE	ASI	ı Ası	р цу	8 Lys	s Asr	ıııe	Phe	Leu	90	Thr	Met	Ile	Lys	Let 95		28	38 -
ASII	Arg	116	100	s Ser	. Lys	CCA Pro	Leu	Gly 105	Glu	Lys	Leu	Leu	Glu 110	Met	Ile	33	6
116	ASI	115	116	e Pro	Tyr	CTT Leu	120	Asp	Arg	Arg	Val	Pro 125	Leu	Glu	Glu	38	4
Pile .	130	Inr	ASI	ıııe	Ala	AGT Ser 135	vai	Thr	Val	Asn	Lys 140	Leu	Ile	Ser	Asn	43:	2
145	GIY	GIU	vai	Glu	Arg 150	AAA Lys	Lys	Gly	Ile	Phe 155	Ala	Asn	Leu	Ile	Ile 160	480	0
Pne	GIY	Pro	GIY	165	Val	TTA Leu	Asn	Glu	170	Glu	Thr	Ile	qeA	Ile 175	Gly	528	
116	GIN	Asn	180	Pne	Ala	TCA Ser	Arg	185	GlY	Phe	Gly	Gly	Ile 190	Met	Gln	576	5
ATG Met	AAG Lys	TTT Phe 195	TGC Cys	CCA Pro	GAA Glu	TAT Tyr	GTA Val 200	AGC Ser	GTA Val	TTT . Phe .	Asn .	AAT Asn 205	GTT Val	CAA Gln	GAA Glu	624	,
Asn.	AAA Lys 210	GGC Gly	GCA Ala	AGT Ser	ATA Ile	TTT . Phe . 215	AAT Asn	AGA Arg	CGT (Arg	Gly :	TAT Tyr 220	TTT :	TCA (Ser /	GAT Asp	CCA Pro	672	!
GCC f Ala 1 225	TTG Leu	ATA Ile	TTA Leu	ATG Met	CAT His 230	GAA (Glu)	CTT . Leu	ATA (His '	GTT 1 Val I 235	TTA (Leu 1	CAT (GGA :	Leu	TAT Tyr 240	720	

GGC Gly	ATT	AAA Lys	GTA Val	GAT Asp 245	Asp	TTA Leu	CCA Pro	ATT	GTA Val 250	Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT	768
TTT Phe	Met	CAA Gln	TCT Ser 260	Inr	GAT Asp	GCT Ala	ATA Ile	CAG Gln 265	Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	Thr	TTT	816
GGA Gly	GGA Gly	CAA Gln 275	Asp	CCC Pro	AGC Ser	ATC Ile	ATA Ile 280	ACT Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC	864
TAT Tyr	GAT Asp 290	Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys 305	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
TCT	GAT A sp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	GAA Glu	TAT Tyr	AGA Arg 410	GGT Gly	CAG Gln	AAT Asn	AAA Lys	GCT Ala 415	ATA Ile	1248
AAT Asn	AAA Lys	Gln	GCT Ala 420	Tyr	GAA Glu	Glu	ATT Ile	Ser	Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	ÁAA Lys	Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ile	GAT Asp	1344
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488
TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	1536

GAT Asp	TTT Phe	: Asr	ı Val	A GAT	GTT Val	CCA Pro	GTA Val	A TAT	GA/	A AAA 1 Lys	A CAI	A CCC	C GC	T AT	A AAA e Lys		1584
AAA	ATT	519 TTT	S ACA	GAT	GAA	AAT	520 ACC) CATO	TTI	CAF	\ TAT	529 TT/	5 A TA(TC	T CAG	•	1632
Lys	530	Phe	Thr	. Ast	Glu	Asn 535	Thr	: Ile	Phe	Glr	540	: Let	ту:	r Se	r Gln		1032
ACA Thr 545	Phe	CCI Pro	CTA Leu	GAT Asp	ATA Ile 550	Arg	GAT Asp	TATA	AGT Ser	TTA Leu 555	Thr	TCI Ser	TC!	TT Pho	GAT Asp 560		1680
GAT Asp	GCA Ala	TTA Leu	. TTA . Leu	TTT Phe 565	Ser	AAC Asn	AAA Lys	GTT Val	TAT Tyr 570	Ser	TTI Phe	TTI Phe	TCI Ser	7 AT0	G GAT Asp		1728
TAT Tyr	ATT Ile	AAA Lys	ACT Thr 580	Ala	AAT Asn	AAA Lys	GTG Val	GTA Val 585	Glu	GCA Ala	GGA Gly	TTA Leu	Phe	Ala	GGT Gly		1776
TGG Trp	GTG Val	AAA Lys 595	CAG Gln	ATA Ile	GTA Val	AAT Asn	GAT Asp 600	TTT Phe	GTA Val	ATC Ile	GAA Glu	GCT Ala 605	AAT Asn	Lys	AGC Ser		1824
AAT Asn	ACT Thr 610	ATG Met	GAT Asp	AAA Lys	ATT Ile	GCA Ala 615	GAT Asp	ATA Ile	TCT Ser	CTA Leu	ATT Ile 620	GTT Val	CCT Pro	TAT	ATA Ile		1872
GGA Gly 625	TTA Leu	GCT Ala	TTA Leu	AAT Asn	GTA Val 630	GGA Gly	AAT Asn	GAA Glu	ACA Thr	GCT Ala 635	AAA Lys	GGA Gly	AAT Asn	TTI Phe	GAA Glu 640		1920
AAT Asn	GCT Ala	TTT Phe	GAG Glu	ATT Ile 645	GCA Ala	GGA Gly	GCC Ala	AGT Ser	ATT Ile 650	CTA Leu	CTA Leu	GAA Glu	TTT	ATA Ile 655	CCA Pro		1968
GAA Glu	CTT Leu	TTA Leu	ATA Ile 660	CCT Pro	GTA Val	GTT Val	GGA Gly	GCC Ala 665	TTT Phe	TTA Leu	TTA Leu	GAA Glu	TCA Ser 670	TAT	ATT		2016
GAC Asp	AAT Asn	AAA Lys 675	AAT Asn	AAA Lys	ATT Ile	ATT Ile	AAA Lys 680	ACA Thr	ATA Ile	GAT Asp	AAT Asn	GCT Ala 685	TTA Leu	ACT Thr	AAA Lys		2064
AGA Arg	AAT Asn 690	Glu	Lys	Trp	Ser	GAT Asp 695	Met	Tyr	Gly	Leu	Ile	Val	GCG Ala	CAA Gln	TGG Trp		2112
CTC Leu 705	TCA Ser	ACA Thr	GTT Val	AAT Asn	ACT Thr 710	CAA Gln	TTT Phe	TAT Tyr	ACA Thr	ATA Ile 715	AAA Lys	GAG Glu	GGA Gly	ATG Met	TAT Tyr 720		2160
AAG Lys	GCT Ala	TTA Leu	AAT Asn	TAT Tyr 725	CAA Gln	GCA Ala	CAA Gln	GCA Ala	TTG Leu 730	GAA Glu	GAA Glu	ATA Ile	ATA Ile	AAA Lys 735	TAC Tyr		2208
AGA Arg	TAT Tyr	AAT Asn	ATA Ile 740	TAT Tyr	TCT Ser	GAA Glu	AAA Lys	GAA Glu 745	AAG Lys	TCA Ser	AAT Asn	ATT Ile	AAC Asn 750	ATC Ile	GAT Asp		2256
TTT Phe	Asn	GAT Asp 755	ATA Ile	AAT Asn	TCT Ser	Lys	CTT Leu 760	AAT Asn	GAG Glu	GGT Gly	ATT Ile	AAC Asn 765	CAA Gln	GCT Ala	ATA Ile		2304
GAT Asp	AAT Asn 770	ATA Ile	AAT Asn	AAT Asn	Phe	ATA . Ile . 775	AAT Asn	GGA Gly	TGT Cys	Ser	GTA Val 780	TCA Ser	TAT Tyr	TTA Leu	ATG Met		2352

AAA Lys 785	AAA Lys	ATG Met	ATT Ile	CCA Pro	TTA Leu 790	GCT Ala	GTA Val	GAA Glu	AAA Lys	TTA Leu 795	CTA Leu	GAC Asp	TTT Phe	GAT Asp	AAT Asn 800	2400
ACT Thr	CTC Leu	AAA Lys	AAA Lys	AAT Asn 805	TTG Leu	TTA Leu	AAT Asn	TAT Tyr	ATA Ile 810	GAT Asp	GAA Glu	AAT Asn	AAA Lys	TTA Leu 815	TAT Tyr	2448
TTG Leu	ATT Ile	GGA Gly	AGT Ser 820	GCA Ala	GAA Glu	TAT Tyr	GAA Glu	AAA Lys 825	TCA Ser	AAA Lys	GTA Val	AAT Asn	AAA Lys 830	TAC Tyr	TTG Leu	2496
AAA Lys	THE	ATT Ile 835	ATG Met	CCG Pro	TTT Phe	GAT Asp	CTT Leu 840	TCA Ser	ATA Ile	TAT Tyr	ACC Thr	AAT Asn 845	GAT Asp	ACA Thr	ATA Ile	2544
CTA Leu	ATA Ile 850	GAA Glu	ATG Met	TTT Phe	Asn	AAA Lys 855	TAT Tyr	AAT Asn	AGC Ser							2574

(2) INFORMATION FOR SEQ ID NO: 22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 858 amino acids
 - (B) TYPE: amino acid

150

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22: Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 140 Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 185

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 215 Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 250 Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 260 Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 280 Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 295 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 310 Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 330 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 345 Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 360 Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 375 Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 390 Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr 425 Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 440 Val Asp`Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys 520 Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln

535

Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 550 Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp 565 Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro 650 Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met 775 Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser 855

- (2) INFORMATION FOR SEQ ID NO: 23:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1644 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..1644

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

ATG Met 1	Pro	GTT Val	ACA Thr	ATA Ile	Asn	AAT Asn	TTT Phe	AA1 Asn	TAT Tyr 10	Asn	GAT Asp	CCT Pro	AT1	GAT Asp	AAT Asn	4	В
AAT Asn	AAT Asn	ATT	ATT E Ile	Met	ATG Met	GAG Glu	CCT Pro	CCA Pro 25	Phe	GCG Ala	AGA Arg	GGT Gly	ACC Thr	Gly	AGA Arg	9	5
TAT Tyr	TAT	AAA Lys 35	Ala	TTT Phe	AAA Lys	ATC Ile	ACA Thr 40	Asp	CGT Arg	ATT	TGG Trp	ATA Ile 45	ATA	CCG Pro	GAA Glu	144	ı
AGA Arg	TAT Tyr 50	ACT Thr	TTT Phe	GGA Gly	TAT Tyr	AAA Lys 55	CCT Pro	GAG Glu	GAT Asp	TTT	AAT Asn 60	AAA Lys	AGT Ser	TCC	GGT Gly	192	?
ATT Ile 65	TTT Phe	AAT Asn	AGA Arg	GAT Asp	GTT Val 70	TGT Cys	GAA Glu	TAT Tyr	TAT	GAT Asp 75	CCA Pro	GAT Asp	TAC Tyr	TTA Leu	AAT Asn 80	240)
ACT Thr	AAT Asn	GAT Asp	AAA Lys	AAG Lys 85	AAT Asn	ATA Ile	TTT Phe	TTA Leu	CAA Gln 90	ACA Thr	ATG Met	ATC Ile	AAG Lys	TTA Leu 95	TTT Phe	288	:
AAT Asn	AGA Arg	ATC Ile	AAA Lys 100	TCA Ser	AAA Lys	CCA Pro	TTG Leu	GGT Gly 105	GAA Glu	AAG Lys	TTA Leu	TTA Leu	GAG Glu 110	ATG Met	ATT Ile	336	
ATA Ile	AAT Asn	GGT Gly 115	ATA Ile	CCT Pro	TAT Tyr	CTT Leu	GGA Gly 120	GAȚ Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	CTC Leu	GAA Glu	GAG Glu	384	
Phe	Asn 130	Thr	Asn	Ile	Ala	Ser 135	Val	Thr	Val	Asn	Lys 140	TTA Leu	Ile	Ser	Asn	432	
CCA Pro 145	GGA Gly	GAA Glu	GTG Val	GAG Glu	CGA Arg 150	AAA Lys	AAA Lys	GGT Gly	ATT Ile	TTC Phe 155	GCA Ala	AAT Asn	TTA Leu	ATA Ile	ATA Ile 160	480	
TTT Phe	GGA Gly	CCT Pro	GGG Gly	CCA Pro 165	GTT Val	TTA Leu	AAT Asn	GAA Glu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	GAT Asp	ATA Ile 175	GGT Gly	528	
ATA Ile	CAA Gln	AAT Asn	CAT His 180	TTT Phe	GCA Ala	TCA Ser	AGG Arg	GAA Glu 185	GGC Gly	TTC Phe	GGG Gly	GGT Gly	ATA Ile 190	ATG Met	CAA Gln	576	
ATG Met	AAG Lys	TTT Phe 195	TGC Cys	CCA Pro	GAA Glu	Tyr	GTA Val 200	AGC Ser	GTA Val	TTT Phe	AAT Asn	AAT Asn 205	GTT Val	CAA Gln	GAA Glu	624	
Asn	AAA Lys 210	GGC Gly	GCA Ala	AGT Ser	Ile	TTT Phe 215	AAT Asn	AGA Arg	CGT Arg	GGA Gly	TAT Tyr 220	TTT Phe	TCA Ser	GAT Asp	CCA Pro	672	

GCC Ala 225	reu	ATA Ile	TTA Leu	ATG Met	CAT His 230	GIu	CTT Leu	ATA Ile	CAT His	GTT Val 235	Leu	CAT His	GGA Gly	TTA Leu	TAT Tyr 240	720
GGC Gly	ATT Ile	AAA Lys	GTA Val	GAT Asp 245	GAT Asp	TTA Leu	CCA Pro	ATT	GTA Val 250	Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT Phe	768
TTT Phe	ATG Met	CAA Gln	TCT Ser 260	ACA Thr	GAT Asp	GCT Ala	ATA Ile	CAG Gln 265	Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	ACA Thr	TTT Phe	816
GGA Gly	GGA Gly	CAA Gln 275	GAT Asp	CCC Pro	AGC Ser	ATC Ile	ATA Ile 280	ACT Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC Ile	864
Tyr	GAT Asp 290	AAA Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys 305	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
AAA Lys	TAT Tyr	AGT	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
ACT Thr	AGA Arg 370	GCT Ala	TCT Ser	TAT Tyr	TTT Phe	AGT Ser 375	GAT Asp	TCC Ser	TTA Leu	CCA Pro	CCA Pro 380	GTA Val	AAA Lys	ATA Ile	AAA Lys	1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400	1200
TCT Ser	GAT Asp	AAA Lys	GAT Asp	ATG Met 405	GAA Glu	AAA Lys	Glu	TAT Tyr	Arg	Gly	Gln	AAT Asn	Lys	Ala	Ile	1248
AAT Asn	AAA Lys	CAA Gln	GCT Ala 420	TAT Tyr	GAA Glu	GAA Glu	ATT Ile	AGC Ser 425	AAG Lys	GAG Glu	CAT His	TTG Leu	GCT Ala 430	GTA Val	TAT Tyr	1296
AAG Lys	ATA Ile	CAA Gln 435	ATG Met	TGT Cys	AAA Lys	AGT Ser	GTT Val 440	AAA Lys	GCT Ala	CCA Pro	GGA Gly	ATA Ile 445	TGT Cys	ATT Ilė	GAT Asp	1344
GTT Val	GAT Asp 450	AAT Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488

TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	15	536
GAT Asp	TTT Phe	AAT Asn 515	GTA Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	TAT Tyr	GAA Glu	AAA Lys	CAA Glņ	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys	15	584
AAA Lys	ATT Ile 530	TTT	ACA Thr	GAT Asp	GAA Glu	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC Tyr	TCT Ser	CAG Gln	16	32
ACA Thr 545																16	44

(2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 548 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15

Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg
20 25 30

Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu

Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 60

Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80

Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe
85 90 95

Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110

Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125

Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 140

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 145 150 155 160

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 170 175

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln
180 185 190

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu 195 200 205

Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro 215 Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 230 235 Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 265 Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 295 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 325 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 360 Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 370 Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 390 Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 410 Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 440 Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 490 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr 505 Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln Thr Phe Pro Leu 545

- 105 -

(2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2616 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

											-					
AT(G CA C Gl:	G TT n Ph	C GT e Va	G AA l As	C AAO n Lys	G CAC	TTO Phe	C AA(2 Asi	TA: 1 Ty:	c Lys	G GA(C CC	T GT.	A AA l As l	C GGT n Gly 5	48
GT: Val	GA L Asi	C AT p Il	T GC e Al 2	u ry.	C ATO	AAA Lys	ATT Ile	CCA Pro	AST	GCC Ala	GG(CAC Gli	G ATO	G1	G CCG n Pro	96
GTC Val	Lys	G GC S Ala	~	C AAC e Lys	G ATT	CAT His	AAC Asn 40	Lys	ATC	TGG	GTI Val	T ATT	Pro	G GA	A CGC u Arg	144
GAT Asp	ACA Thr 50		Γ AC(⊇ Thi	G AAC r Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	Pro	CCG Pro	CCC Pro	G GAA G Glu	192
65	_,	J1 1			70	Ser	ıyr	lyr	` Asp	Ser 75	Thr	Tyr	Leu	Ser	ACA Thr 80	240
			. <i>D</i> ,3	85	ASII	TAC Tyr	Leu	Lys	90	Val	Thr	Lys	Leu	Phe 95	Glu	288
		-7-	100	****	vəħ	CTG Leu	GIY	105	мес	Leu	Leu	Thr	Ser 110	Ile	Val	336
9	O. I	115	110	rne	ΙΙĐ	GGT Gly	120	ser	Thr	Ile	Asp	Thr 125	Glu	Leu	Lys	384
V 41	130	кар	1111	ASII	Cys	ATT Ile 135	Asn	vai	lle	Gln ·	Pro 140	Asp	Gly	Ser	Tyr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	He	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
110	G1 11	riic	GIU	165	гуз	AGC Ser	Pne (GIY :	H15 (Glu '	Val	Leu	Asn	Leu 175	Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT . Ser	ACT (Thr (STII .	TAC A Tyr :	ATT (CGT :	TTC . Phe :	Ser	CCA Pro 190	GAC Asp	TTC Phe	576

ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC' Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glú	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
GCA Ala	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	AAT Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392

469	5		J. U.	Lu A	4°	AT TT sn Ph	ie 11	II A	SII A	SP L	eu A 75	sn I	-ys	Gly	G]	Lu (Glu 480	1440
AT7	T AC	A TO	T GA	AT AC sp Th 48	IT WE	AT AT	A GA e Gl	AA GO Lu Al	CA GO La Al 49	La G	AA G Lu G	AA A lu A	AT sn	ATT Ile	AG Se	r	TTA Leu	1488
GAT Asp	TTI Let	A A1	`A CA e Gl 50	II GI	A TA n Ty	T TA	T TI r Le	A AC u Th	ir Pr	T AA ne As	T T	TT G he A	sp /	AAT Asn 510	GA Gl	A (CCT Pro	1536
GAA Glu	AA1 Asr	AT 1 Il 51	e 3e	A AT	A GA e Gl	A AA' u As:	T CT n Le 52	u Se	A AG	T GA r As	C A	le I	TA (le (25	GGC Gly	CA. Gl:	A I	TA eu	1584
GAA Glu	Leu 530	1.16	G CC t Pr	T AA o As	T AT	A GAZ e Gli 539	1 Ar	A TT g Ph	T CC e Pr	T AA o As	T G0 n G1 54	y L	AA A ys I	AAG Jys	TA:	r G	AG lu	1632
TTA Leu 545	GAT Asp	AA. Ly:	A TA	T AC' r Th	T ATO	G TTO t Phe	CA'	T TA'	T CT	T CG u Ar	g Al	T CA	AA G Ln G	AA lu	TT7 Phe	G	AA lu 60	1680
CAT His	GGT Gly	AA/ Ly:	A TC:	AG0 Ar0 569	3 776	r GCT a Ala	TT)	A ACA	A AA: c Asi 570	n Sei	r GT r Va	T AA 1 As	AC G	lu	GCA Ala 575	L	TA eu	. 1728
TTA Leu	AAT Asn	Pro	AG7 Se1 580	. Arc	r GTT g Val	TAT Tyr	AC#	TTT Phe 585	Phe	TCT Ser	TC.	A GA r As	p T	AT yr 90	GTA Val	L _j	AG Ys	1776
AAA Lys	GTT Val	AAT Asn 595	гÃа	GCI Ala	ACG Thr	GAG Glu	GCA Ala 600	Ala	ATC	TTT Phe	TTI Let	A GG 1 Gl 60	y T	GG (GTA Val	G/ G]	AA Lu	1824
GIII	TTA Leu 610	GTÁ Val	TAT Tyr	GAT Asp	TTT Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	ı Va	A AG 1 Se	GT /	ACT Thr	AC Th	G ir	1872
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	GAT Asp	ATA Ile 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT Tyr	T AT	A GG e Gl	SA C	CT Pro	GC Al 64	a	1920
TTA . Leu .	AAT Asn	ATA Ile	GGT Gly	AAT Asn 645	ATG Met	TTA Leu	TAT Tyr	AAA Lys	Asp	GAT Asp	Phe	· Val	L Gl	y A	CT la 55	TT Le	A u	1968
ATA :	TTT Phe	TCA Ser	GGA Gly 660	GCT Ala	GTT Val	ATT Ile	CTG Leu	TTA Leu 665	GAA Glu	TTT Phe	ATA Ile	CCF	GA G1 67	u I	TT le	GC.	A a	2016
ATA (10	GTA Val 675	TTA Leu	GGT Gly	ACT Thr	TTT Phe	GCA Ala 680	CTT Leu	GTA Val	TCA Ser	TAT Tyr	ATT Ile 685	Al.	GA aA	AT sn	AA(Ly:	3	2064
GTT C Val I	TA i eu 1	ACC Thr	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	AAA Lys	AG Arg	A A.g A:	AT sn	GA/ Glu	A	2112
AAA T Lys T 705	GG (GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	AAA Lys	TAT Tyr	ATA Ile	GTA Val	ACA Thr 715	AAT Asn	TGG Trp	TT/ Let	A GO	la .	AAC Lys 720	5	2160
GTT A Val A	AT A	ACA Thr	CAG Gln	ATT Ile 725	GAT Asp	CTA . Leu	ATA Ile	Arg	AAA Lys 730	AAA Lys	ATG Met	AAA Lys	GAA Glu	A GC 1 A3 73	la	TTA Leu	i.	2208

GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT	CAG Gln 750	TAT Tyr	AAT Asn	;	2256
CAA Gln	TAT	ACT Thr 755	GAG Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	AAT Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT Ile	GAT Asp	GAT Asp	:	2304
TTA Leu	AGT Ser 770	TCG Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	TCT Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	ATG Met	ATT Ile	AAT Asn	ATA Ile	1	2352
AAT Asn 785	AAA Lys	TTT- Phe	TTG Leu	AAT Asn	CAA Gln 790	TGC Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800		2400
ATC Ile	CCT Pro	TAT Tyr	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GAA Glu	GAT Asp 810	TTT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys	. 2	2448
		TTA Leu														Ź	2496
		GAT Asp 835														3	2544
		TTT Phe														. 2	92
		ACT Thr					TAA *									2	616

- (2) INFORMATION FOR SEQ ID NO: 26:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 872 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly
1 5 10 15

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu
50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 135 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 170 Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 215 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu 250 Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 295 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 310 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 345 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 360 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 390 395 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys 440 Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe 450

Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 485 490 Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 535 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu 555 His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 570 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 585 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 600 Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr 615 Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 665 Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp 760 Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys 805

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 825

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 835 840

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 860

Thr Phe Thr Glu Tyr Ile Lys . 870

- (2) INFORMATION FOR SEQ ID NO: 27:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2574 base pairs

 - (B) TYPE: nucleic acid (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

60	A CAACATCATC	A TCGACAACA	C AACGACCCG	CTTCAACTA	CCATCAACA	ATGCCGGTT
120	T CAAGATCACC	C ACAAGGCTT	C GGTCGTTAC	ACGTGGTAC	CGCCGTTCG	ATGATGGAA
180	A GGACTTCAAC	ACAAACCTG	C ACCTTCGGT	GGAACGTTA	GGATCATCC	GACCGTATCT
240	A TTATCTGAAT	ATGATCCAG	C TGCGAGTAC1	TCGTGACGT	GGATTTTCA	AAGAGTAGCO
300	A TAGAATCAAA	AGTTATTTA	G ACTATGATCA	ATTCCTTCAC	AGAAGAACAT	ACCAACGATA
360	TTATCTTGGA	ATGGTATACO	G ATGATTATAA	GTTATTAGAC	TGGGTGAAAA	TCAAAACCAT
420	TGTTAATAAA	CTAGTGTAAC	ACAAACATTG	AGAGTTTAAC	TTCCACTCGA	GATAGACGTG
480	TTTAATAATA	TTTTCGCAAA	AAAAAAGGTA	AGTGGAGCGA	ATCCAGGAGA	TTAATCAGTA
540	ACAAAATCAT	ATATAGGTAT	GAGACTATAG	AAATGAAAAT	GGCCAGTTTT	TTTGGACCTG
600	AGAATATGTA	AGTTTTGCCC	ATGCAAATGA	CGGGGGTATA	GGGAAGGCTT	TTTGCATCAA
660	ACGTGGATAT	TATTTAATAG	GGCGCAAGTA	AGAAAACAAA	ATAATGTTCA	AGCGTATTTA
720	TGGATTATAT	ATGTTTTACA	GAACTTATAC	ATTAATGCAT	CAGCCTTGAT	TTTTCAGATC
780	TATGCAATCT	AAAAATTTTT	CCAAATGAAA	ACCAATTGTA	TAGATGATTT	GGCATTAAAG
840	CAGCATCATA	GACAAGATCC	ACATTTGGAG	AGAACTATAT	TACAGGCAGA	ACAGATGCTA
900	AGGGATAGTT	AAAATTTTAG	AAAGTTTTGC	TATCTATGAT	CGGATAAAAG	ACTCCTTCTA
960	TAATATAT	ACATTAATAT	TCAGATCCTA	AGTTTGCATA	ACAAGGTTTT	GATAGACTTA
1020	ATATAGTATA	CTGAGGGAAA	GTTGAAGATT	ATATAAATTC	TTAAAGATAA	AAAAATAAAT
1080	AGAAACTAAT	TTGGTTTTAC	AGCTTAATGT	ATTATATAAA	GTTTTGATAA	GATGTAGAAA
1140	CTTACCACCA	TTAGTGATTC	GCTTCTTATT	AAAAACTAGA	ATTATAAAAT	ATAGCAGAAA
1200	GTTTAATATA	TAGAGGAAGG	ATCTATACTA	AGATAATGAA	AAAATTTATT	GTAAAAATAA

TCTG.	ATAAAG	ATATGGAAAA	AGAATATAGA	GGTCAGAATA	AAGCTATAAA	TAAACAAGCT	1260
TATG	AAGAAA	TTAGCAAGGA	GCATTTGGCT	GTATATAAGA	TACAAATGTG	TAAAAGTGTT	1320
AAAG	CTCCAG	GAATATGTAT	TGATGTTGAT	AATGAAGATT	TGTTCTTTAT	AGCTGATAAA	1380
AATA	GTTTTT	CAGATGATTT	ATCTAAAAAC	GAAAGAATAG	AATATAATAC	ACAGAGTAAT	1440
TATA	TAGAAA	ATGACTTCCC	TATAAATGAA	TTAATTTTAG	ATACTGATTT	AATAAGTAAA	1500
ATAG	AATTAC	CAAGTGAAAA	TACAGAATCA	CTTACTGATT	TTAATGTAGA	TGTTCCAGTA	1560
TATG	AAAAAC	AACCCGCTAT	AAAAAAATT	TTTACAGATG	AAAATACCAT	CTTTCAATAT	1620
TTAT	ACTCTC	AGACATTTCC	TCTAGATATA	AGAGATATAA	GTTTAACATC	TTCATTTGAT	1680
GATG	CATTAT	TATTTTCTAA	CAAAGTTTAT	TCATTTTTT	CTATGGATTA	TATTAAAACT	1740
GCTA	ATAAAG	TGGTAGAAGC	AGGATTATTT	GCAGGTTGGG	TGAAACAGAT	AGTAAATGAT	1800
TTTG	FAATCG	AAGCTAATAA	AAGCAATACT	ATGGATAAAA	TTGCAGATAT	ATCTCTAATT	1860
GTTC	ATATT	TAGGATTAGC	TTTAAATGTA	GGAAATGAAA	CAGCTAAAGG	AAATTTTGAA	1920
AATG	CTTTTG	AGATTGCAGG	AGCCAGTATT	CTACTAGAAT	TTATACCAGA	ACTTTTAATA	1980
CCTGT	AGTTG	GAGCCTTTTT	ATTAGAATCA	TATATTGACA	АТАААААТАА	AATTATTAAA	2040
ACAA7	TAGATA	ATGCTTTAAC	TAAAAGAAAT	GAAAAATGGA	GTGATATGTA	CGGATTAATA	2100
GTAG	GCAAT	GGCTCTCAAC	AGTTAATACT	CAATTTTATA	CAATAAAAGA	GGGAATGTAT	2160
AAGG	AATTTA	ATTATCAAGC	ACAAGCATTG	GAAGAAATAA	TAAAATACAG	ATATAATATA	2220
TATTO	CTGAAA	AAGAAAAGTC	AAATATTAAC	ATCGATTTTA	ATGATATAAA	TTCTAAACTT	2280
AATGA	AGGGTA	TTAACCAAGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	2340
TCATA	AATTTAA	TGAAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
ACTCI	CAAAA	AAAATTTGTT	AAATTATAA	GATGAAAATA	AATTATATTT	GATTGGAAGT	2460
GCAGA	ATATG	AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
TCAAT	ATATA	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

(2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2574 base pairs
 (B) TYPE: nucleic acid

 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

60	TAATATTATT	TTGATAATAA	AATGATCCTA	TTTTAATTAT	CAATAAATAA	ATGCCAGTTA
120	TAAAATCACA	ATAAAGCTTT	GGGAGATATT	GAGAGGTACG	CTCCATTTGC	ATGATGGAGC
180	GGATTTTAAT	ATAAACCTGA	ACTTTTGGAT	GGAAAGATAT	GGATAATACC	GATCGTATTT
240	TTACTTAAAT	ATGATCCAGA	TGTGAATATT	TAGAGATGTT	GTATTTTTAA	AAAAGTTCCG

ACTAATGATA AAAAGAATAT ATTTTTACAA ACAATGATCA AGTTATTTAA TAGAATCAAA	300
TCAAAACCAT TGGGTGAAAA GTTATTAGAG ATGATTATAA ATGGTATACC TTATCTTGGA	360
GATAGACGTG TTCCACTCGA AGAGTTTAAC ACAAACATTG CTAGTGTAAC TGTTAATAAA	420
TTAATCAGTA ATCCAGGAGA AGTGGAGCGA AAAAAAGGTA TTTTCGCAAA TTTAATAATA	480
TTTGGACCTG GGCCAGTTTT AAATGAAAAT GAGACTATAG ATATAGGTAT ACAAAATCAT	540
TTTGCATCAA GGGAAGGCTT CGGGGGTATA ATGCAAATGA AGTTTTGCCC AGAATATGTA	600
AGCGTATTTA ATAATGTTCA AGAAAACAAA GGCGCAAGTA TATTTAATAG ACGTGGATAT	660
TTTTCAGATC CAGCCTTGAT ATTAATGCAT GAACTCATCC ACGTCCTCCA CGGTCTCTAC	720
GGTATCAAAG TAGACGACCT CCCGATCGTC CCGAACGAAA AAAAATTCTT CATGCAGAGC	780
ACCGACGCAA TCCAGGCAGA AGAACTCTAC ACCTTCGGTG GTCAGGACCC GAGCATCATC	840
ACCCCGAGCA CCGACAAAAG CATCTACGAC AAAGTCCTCC AGAACTTCCG TGGTATCGTC	900
GACCGTCTCA ACAAAGTCCT CGTCTGCATC AGCGACCCGA ACATCAACAT CAACATCTAC	960
AAAAACAAAT TCAAAGACAA ATACAAATTC GTCGAAGACA GCGAAGGTAA ATACAGCATC	1020
GACGTCGAGA GCTTCGACAA ACTCTACAAA AGCCTCATGT TCGGTTTCAC CGAAACCAAC	1080
ATCGCAGAAA ACTACAAAAT CAAAACCCGT GCAAGCTACT TCAGCGACAG CCTCCCGCCG	1140
GTCAAAATCA AAAACCTCCT CGACAACGAA ATCTACACCA TCGAAGAAGG TTTCAACATC	1200
AGCGACAAAG ACATGGAAAA AGAATACCGT GGTCAGAACA AAGCAATCAA CAAACAAGCT	1260
TACGAAGAAA TCAGCAAAGA ACACCTCGCA GTCTACAAAA TCCAGATGTG CAAAAGCGTC	1320
AAAGCACCGG GTATCTGCAT CGACGTTGAC AACGAAGACC TCTTCTTCAT CGCAGACAAA	1380
AACAGCTTCA GCGACGACCT CAGCAAAAAC GAACGTATCG AATACAACAC CCAGAGCAAC	1440
TACATCGAAA ACGACTTCCC GATCAACGAA CTCATCCTCG ACACCGACCT CATCAGCAAA	1500
ATCGAACTCC CGAGCGAAAA CACCGAAAGC CTCACCGACT TCAACGTTGA CGTCCCGGTC	1560
TACGAAAAC AGCCGGCAAT CAAAAAAATC TTCACCGACG AAAACACCAT CTTCCAGTAC	1620
CTCTACAGCC AGACCTTCCC GCTAGATATA AGAGATATAA GTTTAACATC TTCATTTGAT	1680
GATGCATTAT TATTTTCTAA CAAAGTTTAT TCATTTTTTT CTATGGATTA TATTAAAACT	1740
GCTAATAAAG TGGTAGAAGC AGGATTATTT GCAGGTTGGG TGAAACAGAT AGTAAATGAT	1800
TTTGTAATCG AAGCTAATAA AAGCAATACT ATGGATAAAA TTGCAGATAT ATCTCTAATT	1860
GTTCCTTATA TAGGATTAGC TTTAAATGTA GGAAATGAAA CAGCTAAAGG AAATTTTGAA	1920
AATGCTTTTG AGATTGCAGG AGCCAGTATT CTACTAGAAT TTATACCAGA ACTTTTAATA	1980
CCTGTAGTTG GAGCCTTTTT ATTAGAATCA TATATTGACA ATAAAAATAA AATTATTAAA	2040
ACAATAGATA ATGCTTTAAC TAAAAGAAAT GAAAAATGGA GTGATATGTA CGGATTAATA	2100
GTAGCGCAAT GGCTCTCAAC AGTTAATACT CAATTTTATA CAATAAAAGA GGGAATGTAT	2160
AAGGCTTTAA ATTATCAAGC ACAAGCATTG GAAGAAATAA TAAAATACAG ATATAATATA	2220
TATTCTGAAA AAGAAAAGTC AAATATTAAC ATCGATTTTA ATGATATAAA TTCTAAACTT	2280

- 114 -

AATGAGGGTA	TTAACCAAGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	2340
TCATATTTAA	TGAAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
ACTCTCAAAA	AAAATTTGTT	AAATTATATA	GATGAAAATA	AATTATATTT	GATTGGAAGT	2460
GCAGAATATG	AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
CAATATATA	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

CLAIMS

- 1. A polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis, and wherein said second domain is adapted (i) to translocate the polypeptide into a cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into a cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of clostridial neurotoxin precursor that can be converted into toxin by proteolytic action.
- 2. A polypeptide according to Claim 1 wherein said first domain comprises a clostridial toxin light chain.
- 3. A polypeptide according to Claim 1 wherein said first domain comprises a fragment or variant of a clostridial toxin light chain.
- 4. A polypeptide according to Claim 2 or 3 wherein the clostridial toxin is a botulinum toxin.
- 5. A polypeptide according to any preceding claim wherein the first domain exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin.
- 6. A polypeptide according to any preceding claim wherein said second domain comprises a clostridial toxin heavy chain H_N portion.
- 7. A polypeptide according to any of Claims 1-5 wherein said second domain comprises a fragment or variant of a clostridial toxin heavy chain H_N portion.
- 8. A polypeptide according to Claim 6 or 7 wherein the clostridial toxin is a

botulinum toxin.

- 9. A polypeptide according to any of Claims 1-8 further comprising a third domain adapted for binding of the polypeptide to a cell, by binding of the third domain directly to a cell or by binding of the third domain to a ligand or to ligands that bind to a cell.
- 10. A polypeptide according to Claim 9 wherein said third domain is for binding the polypeptide to an immunoglobulin.
- 11. A polypeptide according to Claim 10 wherein said third domain is a tandem repeat synthetic IgG binding domain derived from domain β of Staphylococcal protein A.
- 12. A polypeptide according to Claim 9 wherein said third domain comprises an amino acid sequence that binds to a cell surface receptor.
- 13. A polypeptide according to Claim 12 wherein said third domain is insulin-like growth factor-1 (IGF-1).
- 14. A polypeptide according to any preceding claim comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and a portion designated H_{N} of a botulinum toxin heavy chain.
- 15. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type A.
- 16. A polypeptide according to Claim 15 wherein the botulinum toxin type A light chain variant has at residue 2 a glutamate, at residue 26 a lysine and at residue 27 a tyrosine.

- 17. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type B.
- 18. A polypeptide according to any of Claims 1-13 comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and at least 100 N-terminal amino acids of a botulinum toxin heavy chain.
- 19. A polypeptide according to Claim 18 comprising a botulinum toxin type B light chain, or a fragment or variant thereof, and 107 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 20. A polypeptide according to Claim 15 or 16 comprising at least 423 of the N-terminal amino acids of botulinum toxin type A heavy chain.
- 21. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 22. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain variant wherein residue 2 is a glutamate, residue 26 is a lysine and residue 27 is a tyrosine, and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 23. A polypeptide according to Claim 17 comprising at least 417 of the N-terminal amino acids of botulinum toxin type B heavy chain.
- 24. A polypeptide according to Claim 23 comprising a botulinum toxin type B light chain and 417 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 25. A polypeptide according to any of Claims 14-24 lacking a portion designated

H_c of a botulinum toxin heavy chain.

- 26. A polypeptide comprising a botulinum toxin light chain and a fragment of a botulinum toxin heavy chain, said fragment being not capable of binding to cell surface receptors.
- 27. A polypeptide according to Claim 26 lacking an intact portion designated $H_{\rm c}$ of a botulinum toxin heavy chain.
- 28. A polypeptide according to any preceding claim comprising a variant of a clostridial toxin and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin.
- 29. A polypeptide according to Claim 28 comprising a variant of a clostridial toxin light chain and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin light chain.
- 30. A polypeptide according to Claim 28 or 29 comprising a variant of a clostridial toxin heavy chain H_N portion and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin heavy chain H_N portion.
- 31. A polypeptide according to Claim 28, 29 or 30 obtainable by modification of a DNA encoding the polypeptide so as to introduce one or more nucleotides coding for the cleavage site.
- 32. A fusion protein comprising a fusion of (a) a polypeptide according to any of Claims 1-31 with (b) a second polypeptide being a polypeptide or oligopeptide adapted for binding to an affinity matrix so as to enable purification of the fusion protein using said matrix.
- 33. A fusion protein according to Claim 32 wherein said second polypeptide is

adapted to bind to a chromatography column, such as an affinity matrix of glutathione Sepharose.

- 34. A fusion protein according to Claim 32 or 33 wherein a specific protease cleavage site is incorporated between the first and second polypeptides, said protease site enabling proteolytic separation of first and second polypeptides.
- 35. A composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the botulinum toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*.
- 36. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a positive control in a toxin assay.
- 37. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a vaccine against clostridial toxin.
- 38. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for *in vivo* use.
- 39. A pharmaceutical composition comprising a composition according to Claim 35, a polypeptide according to any of claims 1-31 or a fusion protein according to Claim 32, 33 or 34, in combination with a pharmaceutically acceptable carrier.
- 40. A nucleic acid encoding a polypeptide or a fusion protein according to any of Claims 1-34.
- 41. A nucleic acid encoding a polypeptide or a fusion protein according to Claim

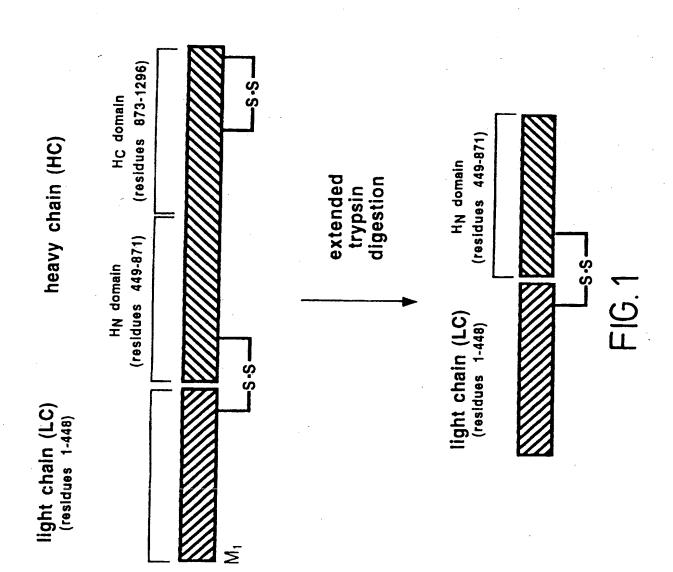
40 and comprising nucleotides encoding residues 1-448 of a botulinum toxin type A light chain.

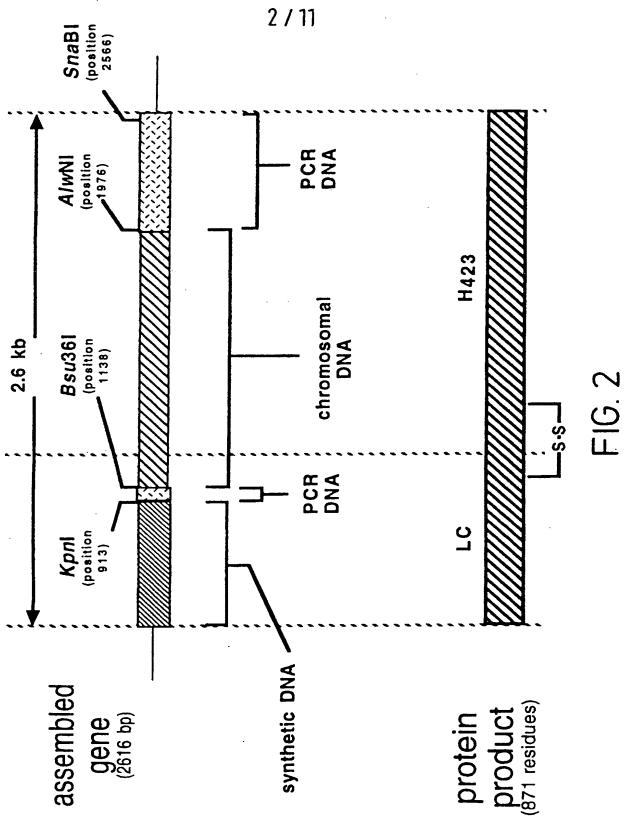
- 42. A nucleic acid according to Claim 40 or 41 comprising nucleotides encoding residues 1-423 of a botulinum toxin type A heavy chain $H_{\rm N}$ domain.
- 43. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 and comprising nucleotides encoding residues 1-470 of a botulinum toxin type B light chain.
- 44. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 or 43 comprising nucleotides encoding residues 1-417 of a botulinum toxin type B heavy chain H_N domain.
- 45. A nucleic acid according to any of Claims 40-44 comprising nucleotides encoding a restriction endonuclease cleavage site not present in native clostridial toxin sequence.
- 46. A nucleotide according to Claim 45 obtainable by modification of a nucleotide encoding a polypeptide or fusion protein according to any of claims 1-34 so as to introduce said cleavage site.
- 47. A DNA according to any of claims 40-46.
- 48. A DNA selected from SEQ ID No:s 1, 8, 10, 12, 14, 16, 18, 23 and 24.
- 49. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid according to any of Claims 40-48 and recovering the polypeptide.
- 50. A method of manufacture of a polypeptide according to any of Claims 1-31 comprising expressing in a host cell a nucleic acid encoding a fusion protein

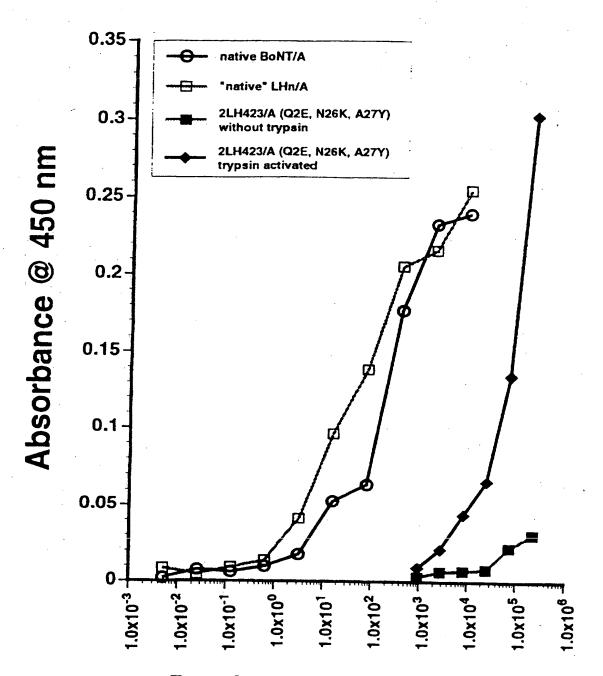
according to Claim 32, 33 or 34, purifying the fusion protein by eluting the fusion protein through an affinity matrix adapted to retain the fusion protein and eluting through said matrix a ligand adapted to displace the fusion protein, and recovering the fusion protein.

- 51. A method of manufacture according to Claims 49 or 50 in which the nucleic acid is DNA.
- 52. A cell expressing a polypeptide or fusion protein according to any of Claims 1-34.









Protein concentration (ng/ml)

FIG. 3

			4/1	1	
2		₹		·	
<u></u>		Ö	<u>.</u>		
<u> </u>		<u> </u>	<u> </u>		
N A G Q M Q P V (Seq I.D. 2)		P V (Seq I.D. 4)	Q P V (Seq I.D. 6)		
>		>	>	>	>
۵		۵.	٥.	٩	٥
0		ď	0	0	0
Σ		⊙ ⁄	∑	∑	∑
0		0	0	0	, o
<u> </u>		<u>ئ</u> ≻	<u> </u>	5	5
A		>	≻	∀	Z Z
<u>-</u>		م	<u>-</u>	2	
A H		H	H	⊢	д Н
×		¥	\checkmark	×	×
H		IAYI	H	H	н
>		>	>	>-	> -
∢		∢	⋖	¥	⋖
H	٥.	H	H	H	H
0	-	Δ .	0 /	۵ /	0 /
> 9	m m	> 5	> 9	>	> 9
z	ш	z	z	ن ح	z
>		>	>	>	>
م	_	٩	٥	۵	٠ هـ
Ω .	0 K L L	۵	0	۵	0
¥		¥	×	¥	¥
>	м П	>	>	>	>-
z	∝	22 11	Z	Z L	Z
ш. ~	40				
¥	⊢	×	×	¥	×
z	∑	Z	z	Z	Z
>	I	>	>	>	>
u.	н	<u>u</u>	ட	<u>u</u>	ш_
MQFVNKQ	G	ш	ш	1 Q F V N K Q	M P F V N K Q
∑ ⊷	GSPGIHMTS	G T M E F V N K Q	G S M E F V N K Q 123	∑ ⊢	∑ ⊣
	ν • • • • • • • • • • • • • • • • • • •	⊢	2 2	,	
	07		– 1		
		_	$\widehat{}$	8	
		<u>Y</u>	Z7.Y	. 199	<
		₹	Α,	12 /B	A, 62
Α/8	*	26K	'A 126K	NOT TO B O	TNO FLUT 119
LH ₄₂₃ /A	1423	∠ ui	1423 E,	e Br tulir pso	BC tulir st a
⇒	23LH423/A	õ	2LH423/A (Q2E, N26K, A27Y)	Native BoNT/A, C. botulinum 2169 Thompson <i>et al.</i> 1990	Native BoNT/A, C. botulinum 62A Binz et al.1990
	N	_		吸られ	以 Ci 盟

= REGIONS OF NON-IDENTITY WITH THE NATIVE SEQUENCES.

1321/441 TCA TTA GAT AAA GGA TAC AAT AAG agc gct gat ggg GCA TTA AAT GAT TTA TGT ATC AAA S L D K G Y N K S A D G A L N D L C I K Eco47 III

1321/441 TCA TTA GAT AAA GGA TAC AAT AAG atc gaa ggt cgt tgc gat ggg GCA TTA AAT GAT S L D K G Y N K I E G R C D G A L N D Factor Xa protease motif

AGG R ACA T GAC D CAG AGG Ř 66A 6 ATG M CCT P GAG E TCT S 16T GCG A AAG K AGG R GTG V CT6 200 R AGG R CA6 AGT S GAA E AGC S 2617/873 TTT ACT F F T 2677/893 GAT GCT D A 2737/913 GGC TCC G S 2797/933 TGT GAT GAA E GT.G V TAT Y GCT A CGG CTG L 999 9 GAG E ACA T <u>م</u> 20 AAG K GCT A)) (1) AGA R 999 AAG K 900 A 767 C CAA O 767 C AAC N GAG E T T GAT D GAT D ر ر ACG T TAT Y GTG V 2587/863 TAC GTA (Y V L 2647/883 CCG GAG P E 2707/903 GGC TTT G F G F 2767/923 GGT ATC G I

=16.7

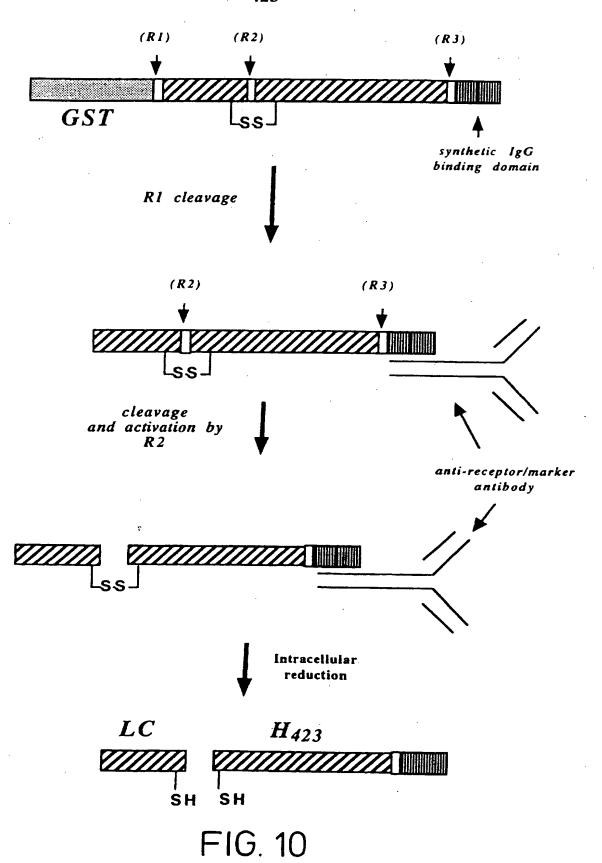
ACA CAT GAT AAG K GAA E 2617/873 TTT ACT (F T 1 2677/893 TAT CAA ACA T T L ATA I T L AGA R CA CA CA CAA O AGA R AAA K GAT D GAT 2587/863 TAC GTA (Y V 1 2647/883 TCT AAA

F16.8

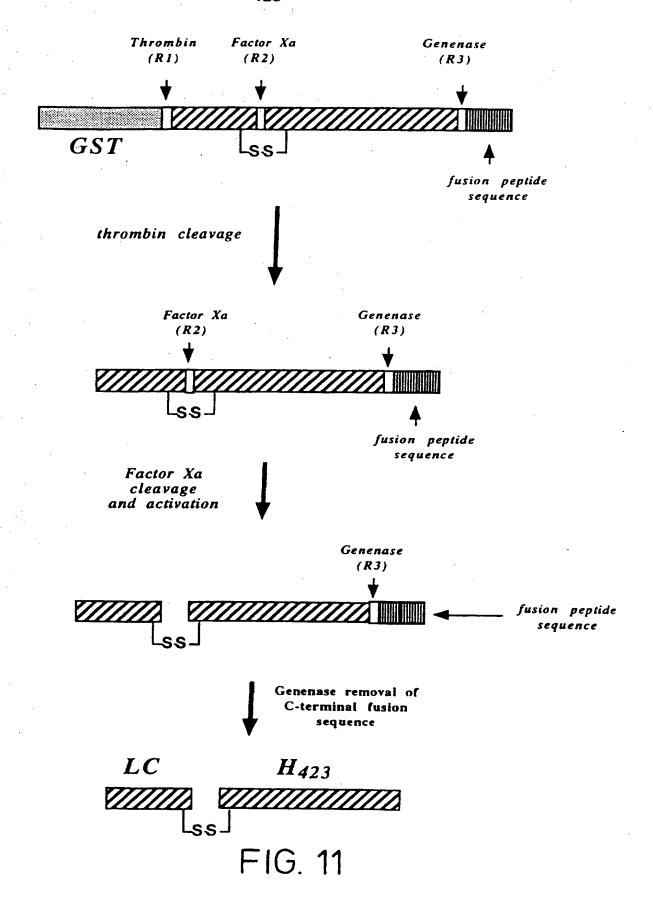
X X CGA R GAA E CAA. GCC AAA K CAA OCA CTG L AAC N GAA E AAC N GCT 666 6 6 770 7 AAA K TTA L GAA E GAA E CGA R S S AAA GCA A GAA E AAC N CAA E _ AAG AAC N AAC AAC N TTC F GAA E TTA L ATT I GAC D TTA L GCT A AAA K GAA E GTA V AAC N AAC N AGC S AAC 2617/873
TTT ACT GAA 1
F T E T
2677/893
GAT GAA GCC GAT GAT TTA CCT H
L P P P
2737/933
CCA AGC CAA I
P S Q S
2857/953
AAA GTA GAC I
K V D R
CCT AAC TTA I
CCT AAC GCT I
A S977/993
CCA AGC GCT I
A S937/1013 TCT ACA
S T
CAA CAC
Q H
ATC TTA
I L
GGT GAC
D D
CCAT TTA
H L
CCA AGC
P S TT L AAA K CAG GCG A GAG E TTA GAC D TTC TAT
F Y
AGT TTA
S L TTA L TAT > GCT A ATC I GAT D AAA CAG AGA R GAT D GAG E T A O CA GCG A CAA TTA L A A N N TAT Y CTA L A GG AAT N AAC N ATC I TT F AGT S GAT D 2587/863
TAC GTA GAT A
Y V D N
2647/883
TCC CCG GGT G
S P G A
2707/903
GAA CAA CAA A
E Q Q N
2767/923
AAC GCC TTC A
A F I
2827/963
CAA AAC GCG I
Q N A F
CAA AAC GCG I
Q N A F
TTC ATC CAA A
F I Q S
3007/1003
AAG CTA AAT G
K L N D S

F1G. 9

8/11 LH₄₂₃/A



$LH_{423}/A^{9/11}$



10/11

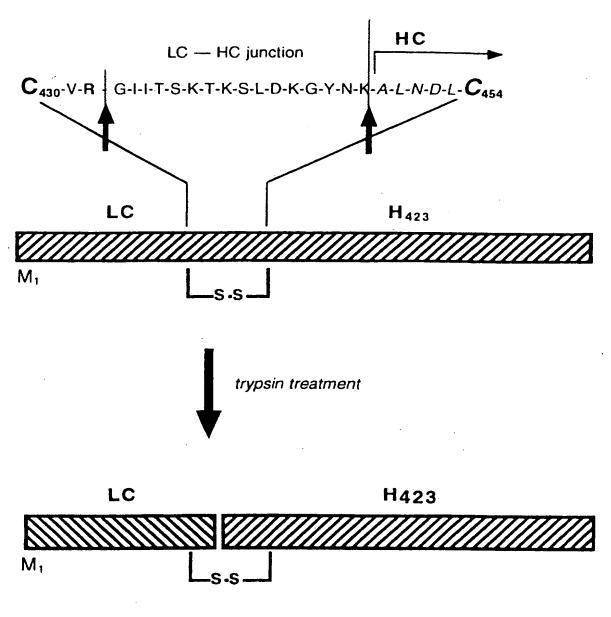
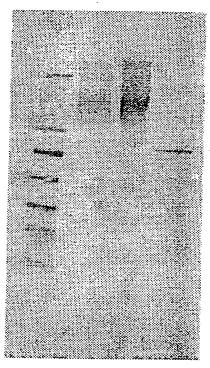


FIG. 12

Panel A.
1 2 3 4



Panel B. 1 2 3 4

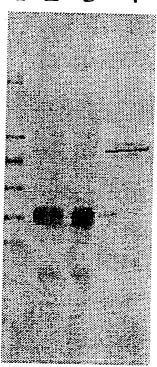


FIG. 13

INTERNATIONAL SEARCH REPORT

unal Application No.

PCT/GB 97/02273 A. CLASSIFICATION OF SUBJECT MATTER IPC 6 C12N15/31 C12N C12N1/21 C12P21/02 C07K14/33 A61K38/16 A61K39/08 According to International Patent Classification (IPC) or to both national classification and IPC **B. FIELDS SEARCHED** Minimum documentation searched (classification system followed by classification symbols) IPC 6 C12N C12P A61K Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practical, search terms used) C. DOCUMENTS CONSIDERED TO BE RELEVANT Category * Citation of document, with indication, where appropriate, of the relevant passages Relevant to claim No. X WO 96 12802 A (OPHIDIAN PHARM INC 1-52 ; WILLIAMS JAMES A (US); PADHYE NISHA V (US); KI) 2 May 1996 see the whole document X KURAZONO H ET AL: "Minimal essential 1-52 *domains* specifying toxicity of the *light* *chains* of tetanus toxin and botulinum neurotoxin type A." J BIOL CHEM, JUL 25 1992, 267 (21) P14721-9, UNITED STATES, XP002047910 see table II -/--X Further documents are listed in the continuation of box C. Х Patent family members are listed in annex. Special categories of cited documents: "T" later document published after the international filing date or priority date and not in conflict with the application but "A" document defining the general state of the art which is not cited to understand the principle or theory underlying the considered to be of partioular relevance invention earlier document but published on or after the international "X" document of particular relevance; the claimed invention filing date cannot be considered novel or cannot be considered to document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such docu-"O" document referring to an oral disclosure, use, exhibition or ments, such combination being obvious to a person skilled in the art. other means *P* document published prior to the international filing date but later than the priority date claimed "&" document member of the same patent family

Date of mailing of the international search report

Hillenbrand, G

Authorized officer

3 0. or 98

Fax: (+31-70) 340-3016 Form PCT/ISA/210 (second sheet) (July 1992)

Name and mailing address of the ISA

Date of the actual completion of the international search

European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo ni,

9 December 1997

INTERNATIONAL SEARCH REPORT

Interconal Application No PCT/GB 97/02273

C.(Continue	Ition) DOCUMENTS CONSIDERED TO BE RELEVANT	
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
x .	LI ET AL: "A SINGLE MUTATION IN THE RECOMBINANT LIGHT CHAIN OF TETANUS TOXIN ABOLISHES ITS PROTEOLYTIC ACTIVITY AND REMOVES THE TOXICITY SEEN AFTER RECONSTITUTION WITH NATIVE HEAVY CHAIN" BIOCHEMISTRY, vol. 33, no. 22, 1994, pages 7014-7020, XP002015938 see the whole document	1
A	BINZ T ET AL: "THE COMPLETE SEQUENCE OF BOTULINUM NEUROTOXIN TYPE A AND COMPARISON WITH OTHER CLOSTRIDIAL NEUROTOXINS" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 265, no. 16, 5 June 1990, pages 9153-9158, XP002009348 see the whole document	1,26,35
	·	
٠	·	
	•	

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inter. .onal Application No
PCT/GB 97/02273

Patent document cited in search report	Publication date	Patent family member(s)			Publication date
WO 9612802 A	02-05-96	US	5601823	Α.	11-02-97
		US	5196193	Α	23-03-93
		AU	3968395	Α	15-05-96
	•	EP	0796326	A	24-09-97
		FI	971732		23-06-97
		МО	971868		24-06-97
		PL	320214		15-09-97
		ZA	9508990		15-05-96
er e		AU	6653894		04-07-94
		CA	2150935		23-06-94
		EP	0671902		20-09-95
	一种 化二氯磺胺二苯		9413264		23-06-94
		US	5466672		14-11-95
		US	5599539		04-02-97
		AU	638786		08-07-93
		AU	6895191		31-05-91
		EP	0498854		19-08-92
		WO	9106306		16-05-91
		US	5443976		22-08-95
		US	5340923	Α	23-08-94

THIS PAGE BLANK (USPTO)

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

perfects in the images include but are not limited to the items checked:	
☐ BLACK BORDERS	
☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES	
☐ FADED TEXT OR DRAWING	
☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING	
☐ SKEWED/SLANTED IMAGES	
☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS	
☐ GRAY SCALE DOCUMENTS	
☐ LÎNES OR MARKS ON ORIGINAL DOCUMENT	
☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY	
OTHER:	_

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)